But What Do They Mean?
Modelling Contrast Between Speakers in Dialogue Signalled by "But"

Kavita Elisheba Thomas

Doctor of Philosophy
Institute for Communicating and Collaborative Systems
School of Informatics
University of Edinburgh
2005

Abstract

Understanding what is being communicated in a dialogue involves determining how it is coherent, that is, how the successive turns in the dialogue are related, what the speakers' intentions, goals, beliefs, and expectations are and how they relate to each other's responses. This thesis aims to address how turns in dialogue are related when one speaker indicates contrast with something in the preceding discourse signalled by "but". Different relations cued by "but" will be distinguished and characterised when they relate material spanning speaker turns and an implementation in a working dialogue system is specified with the aim of enabling a better model of dialogue understanding and achieving more precise response generation.

A large amount of research in discourse addresses coherence in monologue, and much of it focuses on cases in which the coherence relation is explicitly signalled via a cue-phrase or discourse marker (e.g., "on the other hand", "but", et cetera) which provides an explicit cue about the nature of the underlying relation linking the two clauses. However despite research on Speech Acts, planning research into speakers' intentions, and semantic approaches to question-answering dialogues, very little work has focused on coherence relations across turns in dialogue even given the presence of a cue-phrase.

This thesis will explore what sorts of relations the speaker of the "but" perceives between elements in the dialogue, and in particular, it will focus on "but"s communicating *Denial of Expectation*, *Concession*, and *Correction* by determining what underlying cross-turn expectations are denied in the former two, and what is being corrected in the latter case. We will extend work by Lagerwerf (1998) in monologue which presents a treatment for Denial of Expectation and Concession arguing that "but" implicates a defeasible expectation which is then denied (in Denial of Expectation) or argued against (in Concession). We also follow Knott's approach (Knott, 1999a) of describing the semantics of a cue-phrase algorithmically from the agent's mental model of the related utterances.

Task-oriented and nontask-oriented spoken dialogues involving turn-initial "but" are examined, motivating a logical scheme whereby Denial of Expectation, Concession and Correction can be distinguished. These relations are then modelled in the PTT

(Poesio and Traum, 1998) Information State (Matheson, Poesio and Traum, 2000) model of dialogue, enabling more relevant response generation in dialogue systems.

Acknowledgements

A billion thanks go to my supervisor Colin Matheson for all his help along the way. I can't have asked for a nicer, more supportive supervisor in the world. Thanks also to my second supervisor Keith Stenning and to Robin Cooper for great ideas. I want to thank Jon Oberlander for giving me a second chance and letting this thesis happen at all. I spent the last year and a half writing up the thesis in Göteborg, Sweden, and I want to thank Robin Cooper and Elisabet Engdahl for their kind hospitality and also Staffan Larsson for offering implementational help. Thanks to Lori Levin, Alon Lavie, Barbara Di Eugenio and classmates and teachers at Carnegie Mellon and Michael Coen and labmates at the AI Lab, M.I.T. for giving me a start on the strange road of computational linguistics. This thesis also benefitted greatly from the kind help I received from many researchers I met at conferences. In particular I'd like to thank Luuk Lagerwerf and Ali Knott for not getting fed up with my many emails, and also David Traum, Ivana Kruijff-Korbayová, Henk Zeevat and Manfred Stede for helpful comments. I'd like to thank my dissertation committee, Ivana Kruijff-Korbayová, Johanna Moore and also Jean Carletta (with my DDD) for taking the time to read and comment on my thesis. This thesis greatly benefitted from Ivana's extensive list of comments and careful proof-reading and I really appreciate all her help along the way. Thanks to Ben Hutchinson and Caroline Sporleder for lots of good discussions and thanks also to colleagues and friends: David Schlangen, Jurgis Skilters, Mirella Lapata, Stella Psaroudaki and Lori Malatesta. Thanks to Colin Fraser for being a hilarious companion, wise colleague and co-author of Dark Barn, and to Harry Halpin for being the crazy comrade and co-conspirator that he is. Thanks to Chris and Chez and lots of other wonderful punks around Europe and the States that gave me a community and shared amazing ideas. I think not a week went by without frantic emails from my parents asking how things were going and offering support and I want to thank them both more than I can say for being wonderful parents, and for putting up with me and my crazy ideas. Thanks to Leekochama for giving us a home for a few months while this thesis was being written, and thanks to my grandparents, aunts and uncles and cousins and Penny, Phineas and Nicky for being family. Thanks to Scott and mates for late-night horror films and company at gigs and their perverted sense of humour. Thanks to

Emma, Andy, Rob and the Bilstonites for a radical community, and of course to Ulla and Space Bunny and the ACE folk for Actions and solidarity. Thanks to Jacob for being another questioner of things at the beginning of the road. Thanks to old friends Masako, Elicia, Lito and Laura. Thanks to my adopted brothers Hans (a.k.a. “Hokey”, a.k.a. “Count Crab”) and Daniel (“Chico”) and to Maja and Asa for all the love and friendship, ideas and projects. Thanks to the Norder family for a home in Göteborg, and of course, thanks to Micke for being who he is and for loving me. There are lots of other people I want to thank, too many to name here. Suffice to say, the world was kind to me, took me in and nurtured me, and gave me space to grow and I want to let you all know I appreciate it.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(*Kavita Elisheba Thomas*)

I want to dedicate this thesis to my parents, Drs. Hima and T.K. Thomas and to my grandparents, Drs. K.T. and Annie John, who kept telling me they were staying alive just to be able to see me graduate. Unfortunately they are no longer able to travel, but I know they are with me in spirit. They are the reason this thesis finished when it did.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

> "Non-communication is always like the collapse of a revolutionary movement." Raoul Vaneigem, in The Revolution of Everyday Life.

1.1 What Is This All About?

Why is response B1 more reasonable than B2 in the dialogue below?

(1.1)
A: Mary never does her homework.
B1: No, she doesn't.
B2: Apples are good for you.

1.1.1 Coherence

The unacceptability of B2 can generally be described in discourse theory as due to *incoherence*. Understanding what makes a discourse *coherent* is one of the fundamental goals of discourse analysis, since it amounts to understanding what is proposed and communicated by the discourse. Mani *et al.* (1998) define coherence as characterising the deliberate organization of the text by the author in terms of a hierarchical structure to achieve particular argumentative goals. They later add, "[coherence] has to do with macro-level, deliberative structuring of multi-sentence text in terms of relations between sentences and clauses." We will add to their definition by allowing for the

effects of contextual relevance in interpreting dialogue. In other words, we will argue that coherence is not a binary matter, but rather a scale of degree, where context provides the space within which distinctions of coherence can be made.

So considering Example 1.1, B2 becomes more coherent given a context involving dietary requirements, where foods are being evaluated for their merits and Mary not doing her homework is given as an example for her diet being insufficient, and both speakers know that she eats apples. This is an example of macro-level coherence which has to do with "aboutness" (Ginzburg, 1997) of topic more than the more micro-level phenomena like anaphoric reference which depend on theories of local coherence (e.g., Grosz *et al.* (1995); Poesio (1998); Lascarides and Asher (1999)) for resolution.

1.1.2 Discourse Relations

Relations between parts of a discourse are macro-level coherence phenomena, and as Hobbs (1985) points out, an important aspect of understanding a discourse involves recognising the coherence relations in it. Sanders *et al.* (1992) note that coherence relations allow the meaning of the related discourse segments to be more than the sum of the meanings of their parts, indicating that studying coherence relations extends sentential compositional semantics to give additional (noncompositional) semantics.

Determining coherence involves asking what the intended relations between the clauses and sentences of a discourse and their adjacent neighbours are, as well as what relations they share with larger segments of the discourse in which they are embedded. A parallel question to be asked is *which* clause or sentence in the preceding discourse the current clause is related to, and the two questions are interdependent.

1.1.3 Communication Goals

From a cognitive perspective, this involves recognising the speaker's (or writer's) communicative goals at the clausal or sentential level and interpreting their roles in the discourse as a whole. In the case of dialogue, where utterances are frequently not multisentential and often less grammatically correct than in monologue or written text, and where speakers often have different goals, beliefs and expectations, the definition of

coherence needs to also take into account how utterances are related across speaker-turns to determine the coherence of the dialogue as a whole. Interactions between speakers' goals mean that dialogue is a more dynamic, less static form of discourse, which has implications for definitions of coherence that rely on single, overriding intentions that guide the organisation of the discourse, since such views of coherence will need to be modified before they can be applicable to dialogue.

1.2 Thesis Goals

This thesis will attempt to address issues of coherence across speakers and speaker-turns in dialogue. Among other things, this means determining if new utterances fall within the range of what is coherent given the dialogue history. Coherence in dialogue is what accounts for the degree of acceptability of unrelated utterances following each other, as in Example 1.1. In the constrained domain imposed by "but" relating material across turns, this thesis will broadly focus on how speakers achieve their communication goals in these cases.

We will try to determine some of the requirements of coherent dialogue within the specific case of contrast between speakers and across speaker-turns in the hope that by focusing on discourse relations between utterances cued by "but" we can determine more precisely some of the requirements on coherent dialogue. Hobbs (1985) suggests that coherence relations are instantiations in discourse comprehension of more general principles of coherence that we apply in order to make sense out of the world we find ourselves in, so that recognising coherence relations might be just one way in which people simplify their view of the world. In order to avoid ambiguities that might arise if the contrast is not lexically cued, we will focus on the utterance-initial contrastive discourse marker "but" when it relates utterances across speakers and marks a contrast with something in the previous discourse.

1.2.1 Motivation

Statistically "but" appears quite frequently in spontaneous dialogue; for example, in the LDC Switchboard corpus[1] of conversational dialogue, "but" has an average term frequency[2] of 7% and is the 20th most frequent word (out of a total of 41077 words), and is about 25% as likely to occur as the most frequent term. It appears utterance-initially 23.6% of the time when it appears, (excluding cases where it follows parenthetical expressions, false starts or single-word agreements or disagreements like "yeah, but"), and assuming that utterance-initial cases are more likely to involve cross-turn relations, this makes it a fairly statistically significant phenomenon in a large corpus of conversational speech, which provides further motivations for its study.

Currently there is no theoretical account of cross-turn "but" or even the broader cross-turn (cross-speaker) discourse relations (barring question-answering). However the frequency of this word alone as a cue linking material across turns in dialogue is a motivation to investigate how speakers use cues to signal perceived relations between the previous and speaker's turn. There are also broader motivations for studying cross-turn "but"-cued discourse relations; presumably a close analysis of this phenomenon will shed light onto the analysis of other cross-turn cued discourse relations, thereby going a step towards establishing coherence across turns in dialogue.

1.2.2 Thesis Aims

We will focus on language in use in contextualised situations in this thesis, and we will attempt to draw more general characterisations about how the different relations conveyed by "but" might be distinguished. Theoretically we will attempt to determine some of the distinguishing characteristics of the different discourse relations that "but" is involved in in order to present a logical typology of "but"-cued cross-turn relations. To this end, we will present treatments of Denial of Expectation (DofE), Concession and Correction where they relate material across speaker-turns. We will adapt treat-

[1] See http://www.ldc.upenn.edu/cgi-bin/lol/swb for more information and the corpora themselves. Switchboard is a corpus of non-task-based telephone dialogues on randomly assigned topics between strangers.

[2] Term frequency here is simply calculated by normalising the total number of occurrences of a given term by the total number of occurrences of all terms in the corpus, where terms are simply words.

ments of DofE and Concession from monologue which argue that these cases involve implication of a defeasible rule. In DofE this rule is denied by the "but" turn and in Concession both the "but" turn and the preceding turn are related (in different ways) to a contextually relevant claim under discussion. We will then procedurally model the discourse interpretation of "but" in these cases, interpreting what is related and how, and also what relevant expectations, inferences, and beliefs the speaker communicates both explicitly and implicitly. We will then predict how speakers deliberate over what to say next based upon what relations they interpret from the "but" turn.

Among other things, this will involve developing procedures and analytical tools that will aid in making these distinctions, which will hopefully also be applicable in refining characterisations of other discourse relations in dialogue situations. To evaluate our hypotheses in practice, we aim to address coherence in these contrastive cases with a working dialogue model in mind (that is, the PTT (Poesio and Traum, 1998) model of dialogue), specifying both how to implement our findings in this model of dialogue and how this information can be used to predict an appropriate response from the hearer of the "but". This will involve determining what intentions, expectations and obligations the dialogue system should infer, and how a given model of dialogue should update its representation of what is going on in these different situations.

1.3 Thesis Organisation

- We will start by reviewing research on contrastive discourse relations and dialogue and discourse models and present problems of current approaches in dealing with the sort of phenomena we will focus on in Chapter 2.
- We then present motivating examples from the corpora and systematically attempt to determine what is in contrast in order to guide our analysis from the data in Chapter 3.
- We then present a logical scheme whereby we can characterise the different relations studied in Chapter 4.
- In Chapter 5 we present an analysis of Denial of Expectation, one of these

cross-turn “but”-cued relations, model it in a representation of dialogue state and present an interpretation procedure for this relation.

- We analyse and model cross-turn concession and present an interpretation procedure for this relation in Chapter 6.
- In Chapter 7 we analyse and present procedures modelling “but”-cued correction.
- In Chapter 8 we distinguish the relations covered and summarise our analysis and evaluate it, finishing by presenting an implementation of DofE in a working model of dialogue.
- We finish by indicating areas of future research in Chapter 10.

Chapter 2

Literature Review

In order to get at coherence between turns, we will first consider work that addresses coherence between clauses. In this chapter we will start by presenting some contrastive discourse relations which can be signalled by “but” and we will address whether they can be extended to cross-turn relations in dialogue. We will then go on to discuss dialogue modelling approaches, moving on to address problems in extending existing discourse and dialogue approaches to model discourse relations which relate material across turns in dialogue. Finally we will present some necessary additions to dialogue models to enable modelling cross-turn discourse relations.

2.1 Contrastive Discourse Relations

Determining the range of elements that can contrast and the ways in which they can be related will specify the range of acceptable forms of contrast in coherent dialogue. In this section we will explore relations typically associated with “but”, introducing the relations and addressing whether they can be extended to address cross-turn relations in dialogue. There are several possible semantic relations that “but” can impose in discourse:

- Semantic Opposition, (SO, Lakoff (1971))
- Denial of Expectation (DofE, Lakoff (1971))

- Concession
- Corrections or clarifications
- Disagreement

2.1.1 Semantic Opposition

Lagerwerf (1998) distinguishes SO from DofE and concession by claiming that SOs are about two entities in the domain of discussion unlike the other two types of contrast. That is, for SO, while there must be a contrast between the elements compared, it does not have to be a contradiction (e.g., "*Mary* likes *skiing*, but *Anne* likes *chess*"), since if two predicates are about one entity it is difficult for them to be directly opposing without sounding contradictory. For example, consider creating directly opposing predicates by negation, leaving everything else unchanged; then we get something like "Mary likes skiing, but Mary doesn't like skiing", which only makes sense if we take into account prosodic information and search for cooperative interpretations (i.e., the utterance appears to flout Grice's maxim of manner (Grice, 1975)) that could explain why the speaker would utter something that seems contradictory. Following Blakemore (1987), Lagerwerf argues that SO is dependent on *parallel intonation pattern*. He links this to work by Asher (1993) claiming that contrast relations need parallel structure to derive contrast and that parallel structure and intonation are "two sides of the same coin".

However, in dialogue it is possible to have cross-speaker SO without the parallel structure given with two elements in contrast as in the Mary and Anne example above. In fact, in dialogue, it is possible to introduce alternate possibilities that are directly opposed in order to deny a claim, e.g., consider the example below:

(2.1) A: Nobody can do that.
B: But she did it.

It seems that the problem with having directly opposing predicates about the same entity in monologue is not so contradictory in dialogue, since it is more reasonable to have different conversational participants (CPs) with contradictory views or intentions than to have a single speaker with directly opposing views about the same entity.

However in dialogue, directly opposing predicates about the same entity across speakers would probably be categorised as disagreement, e.g.,

(2.2) A: Sue's a vegetarian.
B: (?) But she's not a vegetarian!

In fact, B seems a bit odd. It would be far more natural for B to say "No she's not" if she wants to disagree. This will be discussed more below. However, it is quite reasonable in either dialogue or monologue for the "but" turn to introduce surprising or contradictory information, as seen below:

(2.3) A: Sue's a vegetarian.
B: But she loves hot dogs!

Notice that it seems more acceptable to have the "but" introducing a statement that defeasibly entails a contradiction to A's statement rather than to directly contradict it, (e.g., "She loves hot dogs" defeasibly entails that "She eats meat", which means she is not a vegetarian by definition). This follows from Grice (1975) who argued that "but" shares the logical meaning of conjunction with "and", but carries a *conventional implicature* of contrast. Carlson (1985) argues that only in extreme circumstances does "but" introduce a flat denial, as in Example 2.4 below. He adds that it is relatively more common for "but" to contradict the preceding premise without restating the premise itself, as in Example 2.1 above. He also claims that elliptic dialogue (see Example 2.5 below, B2) conveys an alternative rather than the direct contradiction which can be conveyed in B, and that B2 lacks emphatic force.[1]

(2.4) A: He is dead.
B: But he is not dead!

(2.5) A: He is extremely good.
B1: But he is slow.
B2: But slow.

Another example of "but" introducing direct opposition is categorised as a correction in Kreutel and Matheson (2001b), though it could arguably also function as a disagreement:

[1] Though as Caroline Sporleder pointed out in personal communication, B2 could be read as a concessive use of "but" across speakers as well.

(2.6) A: Helen didn't come to the party.
B: But I'm sure I saw her there.

Here B denies A's claim by providing contrary evidence rather than directly negating A's claim.

2.1.2 Denial of Expectation

Example 2.6 above could possibly also be analysed as DofE, where A's claim leads to the inference that "if Helen didn't come to the party, then no one should have seen her there", which is denied by B. DofE is raised in examples like "Although Greta Garbo was called the yardstick of beauty, she never married." I.e., the asserted proposition that "Greta Garbo was called the yardstick of beauty" implicates the defeasible rule that if someone is beautiful, they will marry, which is then denied in the subsequent clause, (Lagerwerf (1998)).

Related work by Karagjosova (2001) on the German particle, adverb and conjunction *doch* (which as a conjunction is more or less equivalent to "but" in meaning) gives further evidence for the distinction between relations conveyed by "but", since as a sentence adverb and modal or response particle it only has the denied expectation sense. We will argue that her interpretation of *doch* as a modal particle postulates denial of expectation in which it rejects an implication of the other speaker (we will call her A) which counters what the *doch*-speaker (B) believes of A's beliefs. So effectively, according to Karagjosova, B should generate *doch* (as a modal particle) if A expresses something which triggers an implication that goes against what B assumes of A's beliefs. In our approach, we will argue that B can also generate DofE if he himself disagrees with the implication launched from A's utterance. Furthermore, we will model denial of expectation as involving an implication of B's which is denied by contradictory information that he knows. In other words, A's utterance can trigger an implication that B has which is violated by contradictory information which he also possesses. This will be discussed in more detail in the course of the thesis.

2.1.3 Concession

An example of *concession* within the turn (not between speakers) occurs in the following interaction:

(2.7)
A: Shall we take this room?
B: It has a beautiful view, but it is very expensive.

Here the question in *A* gives rise to a *tercium comparationis* (TC, see Lagerwerf (1998)) or claim, which in this case is the offer "let's take this room" which is then argued against concessively by B. B argues for the TC in the clause preceding *but* and argues against the TC in the clause following *but*, but since the second ("but" cued) clause generally indicates the direction of the argument, we interpret *B* above as giving a negative response to *A*'s offer. However, notice that this dialogue involves concession within a speaker turn despite the TC coming from A's question.

In this thesis we will consider concessive dialogues in which material spanning speaker turns contributes to the arguments of the concessive relation. The example below illustrates cross-turn concession signalled by "but", where the TC comes from A's opening claim that Helen didn't attend the party, and the argument for the TC comes from A's 2nd utterance while the argument against the TC comes from B's last (from Kreutel and Matheson (2001b)):

(2.8)
A: Helen did not come to the party.
B: How do you know that?
A: Her car wasn't there.
B: Ok. But she could have come by bicycle.

Lagerwerf considers concession as *epistemic* by definition since it presents arguments for and against the TC, and *asymmetric* since the second clause generally carries the stronger argument, following Spooren (1989). Hobbs (1985) also argues that contrast and violated expectation relations are asymmetric, with the stronger emphasis coming from the second clause; Hobbs' ideas are discussed briefly below. However the asymmetry claim is possibly questionable, and it is particularly unclear whether it would hold across speakers.

2.1.4 Inferential Contrastive Relations: DofE and Concession

Hobbs (1985) classifies the contrast relation as a type of "expansion" relation, where he defines expansion relations as relations that expand the discourse in place rather than carrying it forward or filling in background. He adds that these relations involve recognising or marking inferential relations between segments of text that ease the listener's inference processes. He defines contrast relations as either (1) making contrasting predications of similar entities (where by "similar" he means that they share some reasonably specific property) or (2) making the same predication about contrasting entities.

More specifically, to recognise the first type of contrast he claims that hearers need to infer $p(a)$ from the assertion S_0 and *not* $p(b)$ from the second assertion S_1, where a and b are similar. To recognise the second type of contrast, hearers need to infer $p(a)$ from S_0 and $p(b)$ from S_1, where there is some property q such that $q(a)$ and *not* $q(b)$. His definition of "similar" is quite vague, however, as he himself notes, and he does not point out the possibility that judgements of similarity might be a matter of perspective rather than an arbitrary measure, in which case his definition should mention agents consciously perceiving entities or predications as similar.

He defines another expansion relation (which is quite similar to contrast, especially as Knott sees it (Knott, 1996)) called the "violated expectation" relation, which requires that the hearer infer P from S_0 and *not*P from S_1 in order to be recognised. As an example of violated expectation, he gives the example "John is a lawyer, but he's honest". Hobbs argues that one would draw the inference that John is dishonest from the first clause, which would then get directly contradicted and thus overridden by the second clause. However, in terms of processing these violated expectation relations, it seems that one would need to process the second clause (which is triggered by the contrastive discourse marker "but") in order to infer that there is a contrast between the two clauses, and only then make the inference that John's dishonest from the first clause. This is even more apparent in a subsequent example of his: "This paper is weak, but interesting". Knott (1999a) develops an algorithmic account of these violated expectation relations, which will be discussed subsequently in this thesis.

Hobbs does not mention that in order to infer contrast or violated expectation in

many of these cases, hearers need cues provided either explicitly (i.e., lexically, e.g., by discourse markers) or in the surrounding context in order to make the appropriate inferences and recognise similarities or dissimilarities. He argues that while coherence relations and the discourse markers that signal these relations are not identical, discourse markers impose constraints on the propositional content of the clauses that they modify which are often almost identical to the constraints imposed by the coherence relations themselves. We suggest the objection that since coherence relations require either surrounding context (which could be contained in the related clauses themselves) or discourse markers in order to be recognised, it seems unlikely that the constraints imposed by either can be considered in isolation of the other or be independent. While the semantic relations concession, SO and DofE were all studied in monologue, it remains to be seen whether they involve the same relations across speaker turns in dialogue.

Concession is epistemic, since it provides evidence for and against a TC, and also asymmetric, which in this case means that the second clause carries more force, as discussed by Lagerwerf. We will need to explore the extent to which the argument given in the second clause “wins” in the case of dialogue, and also to what extent the speaker of the “but”-cued clause accepts the other argument. Possibly an explanation for why concession seems strange when it spans speakers arises from the intentional or performative nature of concession as a relation (Mann and Thompson, 1988). That is, it is not simply an informational relation, there is also a strongly argumentative feel about it, in that it intends to concede a minor point and then introduce a much stronger counter-argument, and it gains strength from the minor point it concedes. The strength of the argument gained by conceding an opposing argument gets lost when the conceded point is made by a different speaker, since there is no longer a single intention governing the relation (i.e., both speakers have different intentions). It is not clear whether intentional aspects of discourse relations can be extended across speakers. However if the second speaker, having grounded the conceded point, makes use of it concessively, we have concession across speaker turns. This will be discussed in far greater depth later in the thesis.

2.1.5 Disagreements, Corrections and Clarifications

Unlike the preceding types of contrast, *disagreements*, *corrections* and *clarifications* typically only occur in dialogue. Furthermore in dialogue, it might be the case that contrast can hold with material in the Common Ground as well as with material explicitly present in the preceding dialogue. I.e., if the CPs draw inferences from utterances, subsequent utterances can contrast with these inferences as well as with previous utterances. In fact, as Lascarides and Asher (1999) argue, semantic and pragmatic relations between utterances across speakers are necessary to account for coherence in many cases. Determining coherence is rarely a binary decision and relies heavily on hearers' accommodation processes, so it is only possible to rank degrees of coherence in terms of the ease a speaker in the contextualised situation has in understanding the relation between the current utterance and the dialogue history and conversational context. In some domains (e.g., debate), inferring the semantic/rhetorical relations between utterances might be a necessary condition for understanding the dialogue.

2.2 Dialogue and Discourse Models

Coherence depends on interpretation within context, and context can be seen as providing the space within which to establish coherence. Notions of how structure arises in dialogue led to ideas about dialogue context, for which there are several different approaches, e.g., Grosz and Sidner (1986) and Mann and Thompson (1988). Although the latter theory was formulated with text in mind rather than dialogue, it addresses notions of rhetorical structure which will be relevant for dialogue also. Other models focus on dialogue predominantly, e.g., Information State (IS) frameworks like the TRINDI project (TRINDI (2001), Traum *et al.* (1999)). In this section we will present several approaches to modelling dialogue and discourse in general

While the more traditional discourse structure models like Grosz and Sidner (1986) focus on notions of intentional hierarchies which determine attentional structure that in turn constrains anaphora resolution, the IS update approach to dialogue follows a conversational game paradigm of discourse organisation rather than a tree-structured hierarchy of intentions and determines dialogue moves that are appropriate given a

particular constellation of IS. While intentions can certainly be represented in the IS framework, and are usually accounted for at least minimally, there is also scope for other motivations like social commitments and obligations to be considered in this framework. Utterances in the IS framework are actions that change the information held by the Conversational Participants (CPs). The idea of conversational games specifies appropriate responses (or moves) to utterances in dialogue, and it was partly motivated by work in the conversational analysis tradition (e.g., Goffman (1976), and Sacks and Schegloff (1973)) and also has roots in the BDI (beliefs, desires and intentions) model of Allen and Perrault (1980).

In his book on dialogue games, Carlson (1985) defines coherence in dialogue as determined by whether or not it can be extended into a well-formed dialogue game. He claims that coherence pertains to considerations of dialogue strategy, and defines coherent dialogues as ones in which the moves appropriately serve the dialogue purposes of their authors, linking his definition of dialogue games to notions of goals and intentions. IS update models of dialogue like Traum *et al.* (1999) represent cognitive states of dialogue participants (DPs) via complex data structures with fields specifying the agent's own model of the dialogue history, her model of what she believes the other agent's dialogue model to be, and utterances or beliefs she believes to be mutually accepted, (i.e., grounded utterances that are in the agents' Common Ground). The overhead for such detailed representations of mental state to represent the dialogue model is that plans represented in the IS need to be very fine-grained (Lascarides and Asher (1999)). However, it can probably be argued that as a framework for dialogue modelling the IS update approach allows the greatest representational and implementational flexibility, since the IS is represented simply by an attribute value matrix (AVM), which can be structured differently depending on the sort of dialogue to be modelled and theoretical considerations.

2.2.1 The PTT Model

The PTT model (Poesio and Traum, 1998) is based within the IS framework and addresses determining what actions one should take given a certain constellation of IS, i.e., dialogue update. The IS (as described in Matheson *et al.* (2000)) is partitioned at

the top level into the agent's own model of the dialogue and a field representing what the agent believes to be the other agent's model of the dialogue. This is a reasonable provision, since in order to be engaged in a collaborative dialogue, agents must have some idea of what the other agent's mental model of the dialogue contains in order to make appropriate opening moves and subsequent responses. Furthermore, in order to model disagreements, each CP must have a model of the other CP's IS in their own IS in order to detect the discrepancies in their beliefs that result in the communicated disagreement.

Following Traum and Allen (1994), Matheson *et al.* (2000) shift some of the burden of making appropriate responses from the intentional structure into fields containing socially committed propositions and obligations, both of which are contained in the Common Ground. This is a significant shift away from early models of discourse like Grosz and Sidner (1986) which determine coherence solely on the basis of intentional structure (which is determined by *dominance* and *satisfaction-precedence* relations), toward a notion of social commitments governing one's collaborative behaviour in dialogue. For example, if asked a question, a collaborative and cooperative hearer has a social obligation to try to answer the question, or explain why he cannot answer it otherwise. Furthermore, as discussed in Kreutel and Matheson (1999) and Kreutel and Matheson (2001b), the cooperative hearer has social obligations that require him to respond even when doing so is directly in conflict with his personal intentions.

The dialogue model in PTT also contains a field for grounded information which currently holds grounding acts applying to discourse units (DUs) in the dialogue, and fields to represent the current discourse unit (CDU), previous discourse unit (PDU) and earlier discourse units along with a list of pointers to ungrounded discourse units (UDUs). The CDU, PDU and earlier ungrounded DU fields only hold material to be grounded. All DU fields (e.g., CDU, UDU, PDU, etc) and the grounded field hold obligations, a dialogue history, a list of socially committed propositions, and a list of conditional updates. Obligations, dialogue history elements and conditional updates are all represented via Dialogue Acts, or DAs.

Dialogue Acts in PTT

DAs are the principle unit of information used to characterise DUs; the DA scheme

used in PTT is based on the axiomatisation of DAs presented in Poesio and Traum (1998) which is in turn motivated by the taxonomy of Conversational Acts (CAs) presented in Poesio and Traum (1997). The CA taxonomy was motivated by the discrepancy between theories of context in linguistics (e.g., accounts of anaphora resolution, such as Kamp and Reyle (1993)) and theories of context in planning (e.g, work on intention recognition and dialogue management, such as Traum and Allen (1994)). In the CA taxonomy, core speech acts (SAs) maintain the classical illocutionary acts of SA theory, with the addition of grounding acts, turn-taking acts and higher-level argumentation acts. In the PTT model described in Matheson *et al.* (2000), the inventory of DAs is based on the classification given by the Discourse Resource Initiative (DRI, Discourse Research Initiative (1997)). In Poesio and Traum (1998) the effects of the DRI DAs are specified via a notion of conversational score that keeps track of turn-taking, grounding, intention, and social obligations. The DAs used in Matheson *et al.* (2000) are very basic SAs, and represent acts like asserting, answering, agreeing, acknowledging, directing and requesting information. The effects specified by update rules which characterise these DAs are listed in the latest move field. Given that the DAs are predominantly illocutionary acts, it is probably reasonable that their effects can be predicted even given some minimal context-dependence.

However there is no way in the DRI scheme to characterise more fine-grained discourse relations between utterances like the relations characterised in Rhetorical Structure Theory (RST, Mann and Thompson (1988)). Presumably more fine-grained semantic relations would require more highly context-dependent updates, since semantic relations are much more specifically and uniquely characteristic of the evolving discourse than SAs are, and this would make it much harder to predetermine their effects for all possible contexts.

Information State Model of Dialogue

The IS approach to dialogue modelling characterises the state of each participant's knowledge as the discourse progresses. The work in this thesis follows from the notion of IS put forward by Poesio and Traum (1998) which incorporates a model of grounding (Traum, 1994) and provides a reinterpretation of speech acts (Austin, 1962) and illocutionary force (Searle, 1969). Poesio and Traum's model takes the standpoint that

moves in dialogue involve actions at multiple levels of the discourse structure, e.g., at the level of beliefs, plans, prior discourse history, intentions, social commitments and obligations, and they claim that these actions can be decomposed into Dialogue Acts. Poesio and Traum (1997) propose a modified Compositional Discourse Representation Theory (CDRT) to formalise their notion of IS in a formal semantic framework.

Focusing on the assumption that many basic reasoning processes involved in human interaction can be modelled via IS updates, this thesis proposes update procedures that incrementally operate on certain constellations of information in the IS and result in new ISs that reflect interpretation and deliberation over this information. This follows with the idea that the main effect of utterances is to change the information in the IS in this framework. The actual updating procedure may involve pushing a new value into the components of the IS or merging new and old information. In earlier IS models, dialogue acts were further subclassified as either core speech acts or argumentation acts, (Poesio and Traum, 1998). Argumentation acts (e.g., *correct*, *answer*, *request − evidence*, etc.) characterise context-dependent effects of core speech acts like *assert*, *ask* and *accept*. The relations focused on in this thesis involve updating the IS with predominantly argumentation acts which relate core speech acts to each other. Likewise, the updates we address involve content dependent updates which are triggered by certain IS constellations.

This work follows from research carried out by the TRINDI consortium (TRINDI, 2001). The representation of ISs in the TRINDI project uses feature structures represented as attribute-value matrices which represent the knowledge the dialogue participants have, their utterances and dialogue actions.

The basic aspects of knowledge represented include: a dialogue history (DH) that contains Conversational Acts (CA), private beliefs (BEL), a Task-Plan History (TPH) and Task Beliefs (TB) in Task-Oriented Dialogue (TOD), and conditional rules (COND) which specify what information speakers are socially committed to (Matheson *et al.*, 2000) if they accept this information. The IS is distinguished also by whether information occurred in a current dialogue unit (CDU), a previous one (PDU), or an earlier ungrounded dialogue unit (UDU). Furthermore, different speakers' knowledge can be represented separately by maintaining multiple ISs for each speaker inside the

(1)
$$\begin{bmatrix} \text{PRIVATE} & \begin{bmatrix} \text{BEL [PB1: } bel(A, not(wash(B, dishes))) \text{]} \\ \text{INT [I}_1 \text{ :int(A,ask(B,wash(dishes)))]} \end{bmatrix} \\ \text{UDU} & \begin{bmatrix} \text{DH [CA1: } question(A, why\ didn't\ you\ wash\ the\ dishes) \text{]} \end{bmatrix} \\ \text{PDU} & \begin{bmatrix} \text{DH} \quad \text{[CA2: } answer(B, they\ aren't\ mine) \end{bmatrix} \\ \text{CDU} & \begin{bmatrix} \text{DH} & \text{[CA3: } request(A, wash\ the\ dishes) \text{]} \\ \text{TPH} & \text{[TA1: } wash(dishes) \text{]} \\ \text{TB} & \text{[TB1: } unwashed\ dishes \rightarrow wash\ dishes \text{]} \\ \text{COND} & [accept(B, CA3) \rightarrow scp(B, TA1)] \end{bmatrix} \end{bmatrix}$$

Figure 2.1: Information State Structure in PTT System

IS for the dialogue itself. In this way, aspects of both speakers' knowledge can be represented and compared. Other information which can also be represented include speakers' intentions and obligations.

TRINDI Consortium Systems

There are several different models of dialogue implemented in the IS framework, each of which focus on different aspects of speakers' knowledge and how they come into play when modelling their dialogue. The Cooper-Larsson model (Cooper and Larsson, 1999) is based on the dialogue game board model of Ginzburg (1998) and also includes his Questions Under Discussion (QUD). They focus on how speakers raise issues, which become the current QUD, and how these issues are then resolved. The Poesio-Traum model (Poesio and Traum (1998), Poesio and Traum (1997)) is based on previous research on grounding by Traum (1994) and represents both grounded and ungrounded information. Their model incorporates ideas from the Discourse Research Initiative (1997) and they propose specific updates for locutionary acts, grounding acts, core speech acts, backward-looking and forward-looking dialogue acts. The updates are performed on the evolving CDRS (Compositional Discourse Representation Segment) for the dialogue. Each dialogue act triggers inference rules which specify what other actions are true or must be performed as a consequence of the first one occurring, e.g., if a speaker acknowledges something, then she indicates she understands that. Intentions are also updated and resolved, and there is an action performance rule specifying ability conditions for an agent to perform an action. The main foci of the PTT system are as follows:

- a view of the IS as primarily a record of the dialogue acts that occur during a conversation
- incremental processing, which updates closer to the level of words rather than propositions
- a focus on the social effects of dialogue acts, e.g., commitment to the truth of propositions and obligations to perform actions more than on the more traditional focus on belief and intention which (as argued in TRINDI (2001)) are only deduced via default inferences
- exploration into accessibility conditions of pragmatic processes

The IS consists of the following main parts:

- a private part consisting of beliefs and intentions accessible only to the speaker
- a public part with grounded information (e.g., about which dialogue acts have occurred), obligations and socially committed propositions
- a semi-public part with not yet grounded information divided into dialogue units, e.g., PDU, UDU, CDU

Conversational Representation Theory

Poesio (1995) argues that in order to address frequent dialogue phenomena like fragments constituting utterances, Conversational Representation Theory (CRT) can be utilised, along with the machinery and update of Discourse Representation Theory (DRT), (Kamp and Reyle, 1993). CRT proposes that common ground consists of information about the Discourse Situation (i.e., the set of actions performed by the CPs) rather than on the Described Situation (i.e., the topic of conversation) based on situation semantics (Barwise and Perry, 1983). Poesio claims that utterances introduce *conversational events* which are defined as locutionary acts like *inform, ask, tell, instruct,* and *assert*. Utterance fragments are represented by micro conversational events that the conversational events are built out of, although Poesio (1995) does not address how the conversational events are composed exactly, except to say that doing so

requires AI reasoning and inference rules rather than semantic operations. The common ground contains the micro conversational events, the conversational events, and the subordination relations between them. Poesio (1998) argues that the illocutionary force of an utterance might be a good indicator of whether cross-speaker anaphora can hold or not. For an object or entity in the discourse to be referred to anaphorically, Poesio argues that it must be in the speakers' common ground. However, although CRT predicts accessibility for anaphora across speaker turns, it is unclear how it can help grade degrees of coherence in cases involving phenomena other than anaphora.

Although update in CDRT is monotonic, the monotonicity itself should not be problematic if relations between utterances are added as well as the utterances themselves. A problem with the approach in Poesio and Muskens (1997) is that acknowledgements like "ok" get hard-wired to the conversational event $accept(A,x)$, (where A is the speaker and x is the previous SA). This does not allow the many other uses of "ok", e.g. as an acknowledgement of understanding only rather than acceptance, to be recognised. The authors motivate their approach with the need to account for how a CP's utterance *contributes* to a conversation[2]. However this is still not explicitly accounted for in their approach, since they do not include a way to represent how subsequent utterances are related to previous ones pragmatically or semantically (aside from resolving anaphora).

Poesio and Muskens (1997) argue that the meaning of an utterance in a dialogue is largely determined by the way in which it modifies or updates the conversational score, i.e., the dynamic record of conversational actions in the dialogue. However the DRT accessibility relations cannot predict how the utterances themselves are related rhetorically or semantically unless one utterance explicitly refers to a previous one. For example, reference to SAs is incorporated into CRT, so if B replies to A with "Can you repeat that", A's SA will become the argument of B's repeat act via a $request(repeat)$ conversational event that gets added to the conversational score. For disagreements, this creates a problem, because conversational events are concatenated with the conversational score, as in the example below:

[2] As in Clark's sense of contributing–see his essay "Contributing to Discourse" in Clark (1992).

(2.9) A: George and Bill are exactly alike.
B: But Bill plays football.

According to CRT, A should introduce the conversational event *ce*1: *assert* $(A, B, [x, y, s1 : George(x), Bill(y), s1 : alike(x,y)])$. B's reply should then introduce the event $ce2 : inform(B, A, [y, w, s2 : Bill(y), football(w), s2 : likes(y, w)])$. If these two events were concatenated with the conversational score, they would be potentially contradictory, assuming that from what we know so far in the dialogue, processing B's utterance should lead to the defeasible inference that "George doesn't play football", since the "but" triggers a contrast between the two utterances. That is, assuming the defeasible inference above, we would have the contradictory propositions "George and Bill are exactly alike" and "Bill plays football and George doesn't play football", which contradicts the proposition that they are exactly alike. However in conversation this seems like a perfectly reasonable thing to do, since we are not claiming that George and Bill are identical objects mathematically, so there is nothing to bar the speaker from noting exceptions to the rule, which they signal with "But" in this case. However how would we handle a case where the contrast is not explicitly marked as in "Bill likes football" (without "but") following A? In order to infer indirect SAs, it seems tempting to rely on adjacency pair models (Sacks and Schegloff (1973)) to infer that B's reply must be either an agreement or a disagreement to A's assertion.

2.2.2 Other Dialogue Modelling Efforts

Segmented Discourse Representation Theory

Lascarides and Asher (1999) argue that rhetorical relations are necessary in order to account for anaphora resolution across speaker turns, and incorporate rhetorical relations in their Segmented Discourse Representation Theory (SDRT). They address the necessary accompanying pragmatic reasoning in their DICE theory. SDRT addresses both indirect SAs (Asher and Lascarides, 2001) and discourse relations (Lascarides and Asher, 2002), but does not address cases involving implication like DofE or concession. Only contrast and SO relations are specified, and both have rigid criteria, e.g., parallel structure for contrast (Asher, 1993). Furthermore, there is no account of speaker's expectations or socially committed propositions. While SDRT addresses

dialogue as well (Asher and Lascarides (1998b), Lascarides and Asher (1999), Asher (1998)), it does not address how differences in speakers' plans and beliefs are resolved, both of which are key characteristics of the phenomena we study in this thesis.

Adjacency Pair Model

Sacks and Schegloff (1973) try to address the expectation CPs have of how their opening conversational moves will be responded to by the other CP, so they predict simple pairs of responses like Question-Answer, Assessment-Agreement/Disagreement, etc. They allow for nested pairs (e.g., nested questions, as in Example 2.10 below), and their approach might help to predict that an indirect answer like "Bill likes football" is either an agreement or disagreement with the assessment that A makes in Example 2.9.

(2.10)
A: When's the next bus to Dogville?
B: Do you want to leave from Cattown or Eggsby?
A: I'm leaving from Cattown.
B: Oh, well then the next bus is at 4:30.

However, it will not help resolve the mixed agreement and disagreement in the concession in Example 2.11 below:

(2.11)
A: That fi lm was terrible.
B: But I thought Thora Birch did a fabulous job.

The adjacency pair approach is very rigid and does not account for problems with the question itself as in example 2.12 below, which shows that the question is not a valid one and cannot be answered, either by stating that the question itself is invalid, as in B1, or by stating that the questioning itself is not allowed, as in B2 (neither of which gets noted in the rigid adjacency pair model, which is expecting an answer to the question to close the game).

(2.12)
A: When's the next bus to Dogville?
B1: There is no bus to Dogville.
B2: That's classifi ed information; you're not allowed to know that.

This model similarly cannot account for digressions, interruptions, or misunderstandings. That is, the adjacency pair model does not allow the conversational game being played to be changed or abandoned. If more than one miscommunication occurs, the game fails, as in example 2.13 below:

(2.13)
A: When's the next bus to Dogville?
B: There's a 2.45 from Piddles to Dogville.
A: No I meant leaving from Bogtown to Dogville.
B: Oh, well there's a 3:15 from Bogtown then.

It also does not address how implicit acknowledgement can be conveyed by continuing the discussion, since the continuation is itself an opening move, and there is no explicit closing move (see example 2.14 below).

(2.14)
A: Eggsby is a lovely place to live.
B1: And not too pricey.

TRAINS and the DAMSL and DRI schemes

DAMSL, or Dialogue Act Markup in Several Levels, is an annotation scheme for dialogue acts used in the TRAINS project[3] and is described in Allen and Core (1996). The DAMSL annotation scheme was based on the DRI (Discourse Research Initiative (1997)) dialogue coding standards working group; the goal of the DRI is to develop standards for annotating corpora with semantic, pragmatic and discourse features, and in Discourse Research Initiative (1997) they focus on annotating dialogues. The DRI proposal for dialogue annotation presents a multi-level annotation scheme with *Forward Looking Communicative Functions (FCFs)* that describes "how the current utterance constrains future beliefs and actions of the participants, and affects the discourse" (Allen and Core (1996)), *Backward Looking Communicative Functions (BCFs)* which describe how the current utterance relates to the previous discourse, *Communicative Status (CS)*, which records whether the utterance is intelligible and successfully completed, and *Information Level (IL)*, which characterises the semantic content of the utterance. Further subdivisions within each category are quite simple, and very likely insufficient for more complex and wide-ranging non-task-related conversation. For example, Information Level only characterises whether the utterance is involved in doing the task, task-management (talking about the task) or communication-management (maintaining the communication); there is an additional "other" category to hold anything that does not fit into these categories.

[3]The TRAINS project involves task-oriented problem-solving dialogues; See http://www.cs.rochester.edu/research/trains/

While it is reasonable that communication-management would apply to all types of conversations (not just task-related ones), the other categories are clearly inadequate for non-task-related dialogue which might not even stick to a single topic. Similarly, the FCFs are either statements, information requests, influencing the listener, or committing the speaker to an action; influencing the listener covers all non-self-committing utterances and non-information-requests, which means everything else, clearly too broad a category to distinguish between the sorts of discourse relations we might want to capture as FCFs in non-task-related conversation to represent how utterances are related to each other. Likewise for BCFs, which can indicate agreement, understanding, answering, information-relations or antecedents.

While information-relations do not explicitly address discourse relations between utterances, i.e., nothing is mentioned explicitly about applying them, this would be the logical place in the DRI scheme for incorporating an account of discourse or semantic relations, although doing so would clearly require a restructuring of the information-relation category so that it is not so biased toward task-oriented dialogues (e.g., the task management and task subcategories could presumably be modified for conversational dialogues). For the FCF category, there is also no equivalent category for nucleus-final relations (where by "nucleus-final" we refer to RST terminology, i.e., relations in which the satellite precedes the nucleus); so modelling nucleus-final discourse relations like concession in the FCF category would require some thought.

Moreover, the categories CS, IL, FCF and BCF do not represent orthogonal dimensions, and this might lead to some coding ambiguity in a more complex domain which requires more subtypes, for example, rhetorical relations with the nucleus preceding the satellite might be viewed as BCFs, and vice-versa for FCFs, while they might really belong under the Information Level category.

In the DRI manual they also mention an Information-Status category which does not get mentioned in the DAMSL manual; they propose four categories of information-status: repetition, reformulation, inferences, and new, where the first three categories correspond to "old" information. Since they do not describe the categories in much detail, it is hard to determine whether they will be sufficient for non-task-based dialogue, and they do not really seem clearly motivated. However as we shall see later, char-

acterising the information-status of utterances explicitly in dialogue is crucial to determining whether they are performing grounding or helping the dialogue to progress. And while the DAMSL project includes speech acts and grounding acts, it does not include any sort of discourse relations between utterances, which as we shall see later, are crucial for characterising how utterances are actually related to each other in a dialogue.

Traum and Nakatani's Variation on the DRI Scheme

Traum and Nakatani (1999) propose two schemes based on a principled study of different annotation schemes and on development within the DRI initiative that argues that intentions are reflected at all levels of granularity in dialogue coding but can only be considered as shared at the level of grounding content, and that grounding units should therefore be taken as the starting point for coding intentional structure. They claim that the dialogue act (DA) level of the DRI scheme represents intentions relating to communication rather than task management. They contrast game structure in the MapTask coding scheme (see below) with Common Ground Units (CGUs; see Traum and Nakatani (1999)) which focus on how the content of DUs gets added to the common ground[4] and claim that game structure in MapTask codes achievement of dialogue purposes, while CGUs only cover grounding.

The CGU analysis focuses on establishing what has been said, and proposes higher-level intentional or informational units (IUs) built hierarchically of CGUs which describe informational and intentional relationships between what has been said. They model intentional structure following Grosz and Sidner (1986), and informational structure following earlier work by Pollack involving just two relations, "generates" and "enables"; it would be interesting to investigate whether a richer representation of both structures would enable greater representational capabilities for conversational dialogue, e.g., RST subject-matter relations to represent informational structure.

The IU level of analysis focuses on topic-structure of the dialogue and involves reasoning about lower-level CGUs. In Nakatani and Traum (1999) they discuss coding rhetorical relations at the lowest (or "micro-range") level of the discourse, which corresponds to coding at the utterance level, and raise as a question the utility of coding

[4]Traum and Nakatani (1999) makes an interesting distinction between coherence which relates to the content of DUs and coherence which relates to how content gets added to the CG.

domain/task-independent rhetorical relations, although as we will see, there is certainly a need for this sort of annotation effort.

Modifying RST for Dialogue

Since the TRAINS project is one of the longest running research efforts on spoken dialogue, we will study many of the findings and reports from this project; however, because TRAINS involves only task-oriented dialogues, we will need to investigate what is required to extend the DAMSL annotation scheme to handle the phenomena we will encounter in spontaneous conversational dialogue. For example, in Stent (2000) the hierarchical structure of RST is correlated within task-oriented dialogues, which presumably involve a hierarchical organisation of goals. Stent (2000) goes on to propose a partial decision tree for presentational relations modified for dialogue and she distinguishes between affecting the hearer's beliefs/attitudes/ability to perform some action on the one hand, and trying to increase the hearer's belief or understanding of some fact on the other. She cites arguments from Moore and Pollack (1992) and Moore and Paris (1994) that subject-matter relations should be kept distinct from presentational relations, and if both relations apply, she advocates removing RST's uniqueness condition and labelling the given span with both applicable relations. She also addresses cross-speaker elaboration and sequence relations and incorporates some adjacency-pair relations and schemas in her modification of RST for task-oriented dialogue: e.g., question-response, proposal-accept (adjacency pairs), make-plan and describe-situation (schemas). The extent to which adjacency-pairs and schemas will work in spontaneous conversation is a matter of debate however.

The Collagen Project and Collaborative Discourse Theory

Another approach to dialogue modelling which focuses specifically on collaborative task-oriented dialogues involving humans and machines is the Collagen project (Rich *et al.*, 2000). Collagen is an implementation of Lochbaum's Collaborative Discourse Theory (CDT) (Lochbaum, 1998) which in turn extends SharedPlans (Grosz and Kraus, 1999). SharedPlans presents a model of collaborative planning which also handles cases in which agents have partial knowledge and share responsibilities (as well as contracting out actions). CDT extends SharedPlans via a computational model which recognises intentional structure and utilises it in discourse processing. Sidner (1994)

defines discourse exchanges involving proposing, counter-proposing, rejecting and providing supporting information in Collagen, focusing on the negotiations involved when agents try to reach agreement on the beliefs relevant to their collaboration.

However her artificial language of negotiation does not get at the subtleties involved in either interpreting or generating cues like turn-initial "but" which signal cross-turn relations like DofE or concession. In a simple negotiation dialogue (from the same paper), she has situations like the ones we consider in this thesis, shown below followed by their representations in her artificial agent language; PFA below stands for Proposal to Accept, AP is an Action Proposal and CO counters a belief by offering an alternative belief.

(2.15) C7: Ok. The result of F is W.
(AP C (Should-Do C (Provide-Support C (Contributes (F;B) Z))) R)
(PFA C (Equal (Result F) W) R)
R8: Yeh but A is easier than F, so A;B, then D;E generates Z.
(AP R (Equal (Result F) W) C) (CO R (Generates (((F;B) & D); E) Z))) C (And (EasierThan A F) (Generates (((A;B) & D);E) Z)))
C9: Yeh A's easier, so A's enabling conditions are different from F's so A can't be in the recipe.
(AP C (EasierThan A F) R) (CO C (Generates (((A;B) & D); E) Z) R (Not (Equal (Result A) (Result F))))
R10: But they are the same. I've done A;B.
(CO R (Not (Equal (Result A) (Result F))) C (Equal (Result A) (Result F))) (PFA R (Done R A;B) C)

Notice that countering moves (CO) occur in both "but" turns, R8 and R10, and work very similarly to what we term cross-turn concession. However our focus, as will be seen in Chapter 6, is on the different ways in which cross-turn concession can relate material across turns, and we explicitly focus our interpretation efforts on cases involving turn-initial "but", while the focus here is instead on the types of moves agents can make in negotiating over how to go about achieving a collaborative task, so the particular focus on "but"-cued turns, or more broadly how these agents realise their communication goals (expressed in terms of Sidner's artificial language) is not addressed. As this thesis will argue, "but" can be involved in far more than simply countering by proposing an alternative; we will present DofE and correction, both of which communicate different information and require different interpretations.

The COCONUT Project

Another project of interest includes the COCONUT project (Di Eugenio *et al.* (1998)), a constraint-satisfaction project involving task-oriented dialogues which builds upon the DRI and DAMSL annotation scheme. The COCONUT annotation scheme adds a *topic* feature to the DRI scheme which tries to capture what the utterance is about, and further distinguish this into *topic* and *attitude*. This scheme also incorporates the elaboration rhetorical relation along with a summary schema and some adjacency-pairs like *act:condition* and *act:consequence*. However, they do not keep rhetorical structure separate from other types of dialogue behaviour, a distinction which Stent (2000) argues for. She claims that including rhetorical relations in the DAMSL scheme would enable 40% of utterances in TRAINS dialogues which are sequences of statements to be related, where currently there is no mechanism for explaining how they are related to each other.

The MapTask Project

Carletta *et al.* (1996) describes the HCRC MapTask project's dialogue annotation effort; MapTask is similar to COCONUT and TRAINS in that it involves task-oriented dialogues between humans working together to solve a problem, (which is navigational in this case). Like the DAMSL scheme, the MapTask annotation scheme does not include discourse relations; however it includes adjacency-pairs and follows a conversational game paradigm by describing appropriate move sequences. The coding scheme used to annotate the MapTask focuses on distinguishing between *utterance function*, *game structure* and *higher-level transaction structure*. *Transactions* describe subdialogues, which are made up of *conversational* or *dialogue games*, following Carlson (1985) and others. The idea behind conversational games is that conversational participants (CPs) are involved in games which satisfy their individual goals through joint activity. Games are distinguished by their purpose, e.g., information-getting, task-solving, etc., and are subdivided into dialogue moves that progress the CPs in their games, e.g., *initiating* moves like *instruct, explain, check, align*, etc., *response* moves like *acknowledge, reply-yes, reply-no* and *clarify, ready* moves indicating that the CP is ready to perform some action or get involved in some dialogue game. *Align* is a grounding move, so grounding is not distinguished in dialogue games from other sorts

of dialogue moves. Moves typically cut across the distinctions made in DAMSL between communication-management and task-management, and also do not distinguish between semantic relationships between utterances and pragmatic ones.

Summary of dialogue modelling issues

The main points that arose when considering various dialogue modelling approaches were:

- Agents need to be aware of other agents' mental model of the dialogue in order to be able to respond appropriately; this is especially important in cases of misunderstanding and disagreement, and is required to make corrections, etc.
- None of the approaches we have seen incorporate fine-grained discourse relations between utterances except for SDRT (Lascarides and Asher (1999)), and it is clear that we will need to incorporate some account of discourse relations in order to address "but" and the relations it imposes in dialogue as SDRT does not address relations involving implied expectations or relations w.r.t. a TC.
- The approaches we have seen mostly address task-oriented dialogue, and we will need to investigate what is required to address the phenomena we encounter in nontask-oriented dialogue as well.

2.3 Problems with Existing Work

One hypothesis to be examined is the extent to which recognising the discourse relations between cross-speaker utterances determines perceived coherence of the dialogue. However, while there has been much work on discourse and discourse relations, very little of it has been applied to dialogue relating cross-turn material, and as discussed below, it is not always a straightforward matter to just use an "off-the-shelf" theory of discourse relations in dialogue. In this section we will consider some of the shortcomings of modelling cross-turn relations with existing approaches. In particular, we will examine RST, Grosz and Sidner's discourse model, the PTT approach and Knott's approach to modelling the semantics of "but" and consider how they must be extended to account for cross-turn "but".

2.3.1 Applying RST for Dialogue

Rhetorical Structure Theory (or RST) argues that a property of coherent texts is that they can be organised hierarchically (and compositionally) by *rhetorical relations* holding between consecutive spans of text at the bottom of the tree and between relations above the leaves of the tree. To this end, Mann and Thompson (1988) define a set of rhetorical relations that operate on discourse elements (i.e., clauses, sentences and groups of sentences). In RST, contrast is formulated via "concession", "contrast" or "antithesis" relations.

However, RST was designed to operate on text, not dialogue, and employing RST's contrast relations across speakers means adapting the definition of these relations, since the *satellite* and *nucleus* (or *nucleii* in multinuclear contrast relations) are spoken by different speakers, and can therefore have multiple and possibly contradictory intentions guiding their combination. Mann and Thompson formulated rhetorical relations to be guided by the speaker's (or writer's) intentions in structuring their discourse in a certain way. The two main types of relations in RST, *subject matter* and *presentational* relations are both defined with respect to the speaker's goals; i.e., "subject matter relations are those whose intended effect is that the reader recognises the relation in question" and are linked with semantic relations, while "presentational relations are those whose intended effect is to increase some inclination in the reader", and are linked with pragmatic relations.

Moore and Pollack (1992) and others (e.g., Grosz and Sidner (1986)) have argued that RST does not fully specify the intentional structure of a text, and Grosz and Sidner note that it does not allow for multiple relations to hold simultaneously between consecutive elements in a coherent discourse. Moore and Pollack argue that relations need to hold at both the *informational* and *intentional* levels simultaneously; i.e., subject matter relations like causal relations relating the semantic content of clauses can hold simultaneously with presentational relations guiding the intended effect of the related clauses on a hearer. Hobbs (1996) has also argued for informational and intentional perspectives to be considered in parallel when analysing discourse.

Moore and Pollack argue that recognition of discourse relations may flow from the intentional to the informational level or vice-versa. For example, in their example

"George Bush supports big business. He's sure to veto House Bill 1711", they argue that if one does not know what House Bill 1711 proposes, one might deduce from the conversational context (e.g., a debate or argument, where the CPs have opposing views and the speaker is trying to convince the hearer that Bush will veto more bills) that S_1 is evidence for S_2 (first and second sentences respectively in this example), i.e., the hearer will reason from intentional coherence to informational coherence. They show that the reasoning might go the other way if one knows already what House Bill 1711 proposes; e.g., if House Bill 1711 is against big business, then one might reason that S_1 is a cause of S_2, reasoning from informational coherence to intentional coherence. Since we do not know a priori what the conversational context or background knowledge of the speakers in a dialogue are, and since it is often possible to recognise different rhetorical relations depending on whether one reasons from informational coherence to intentional coherence or vice versa, this argument gives weight to the claim that rhetorical relations as defined in RST are often ambiguous, and not unique, as the RST claim holds.

In terms of dialogue, and in particular, cross-speaker contrast, it seems likely that presentational relations as defined in RST will not be applicable to relating utterances across speakers without extensive redefining, since they have the intended effect of increasing some inclination in the other CP, and so involve a single overriding intention, which implies a single speaker behind it, and we are considering cross-speaker relations here. In RST, concession is defined as presentational, with the overriding intention to increase the reader's regard for the nucleus, as opposed to the subject-matter relation contrast, which simply presents two contrasting arguments without any intentions to increase the reader's regard for either argument.

It is conceivable that a speaker B could semantically relate what they are saying to a previous comment by A with the general expectation that A recognises the relation between what B said and what preceded this in the dialogue history in order to do away with presentational relations in dialogue. However, if one also wants to account for intentions behind responding in the way one does in dialogue, then one will have to consider planning approaches that consider multiple CPs' interactions, like work by Grosz and others on SharedPlans (Grosz and Kraus (1999)) and Clark's work on

Joint Activity (Clark, 1996)[5] since RST was formulated with text (written by a single author), rather than dialogue, in mind. Also it is clear that subject-matter relations alone cannot account for relations involving implications like DofE and concession (as we saw it earlier).

2.3.2 Grosz and Sidner's Discourse Model Structure

Grosz and Sidner (1986)) do not assign semantic relations between units in a text as RST does; they break a discourse up into its *linguistic structure*, *attentional structure* and *intentional structure*. The intentional structure is a hierarchical structure of the speaker's intentions, which are ordered by *dominance* and *satisfaction-precedence* relations (i.e., informally, parent-child relations and left-to-right ordering). In Grosz and Sidner's model, intentional structure determines discourse coherence (or "discourse purpose" in their terms); a discourse is judged coherent if it can be ordered into a single tree of intentions, which reflects the relations between discourse goals in terms of dominance and satisfaction-precedence. Recognising the discourse purposes (and consequently the "discourse segment purposes" of the lower-level nodes) is crucial to understand the discourse. The intentional structure also determines attentional structure, which dynamically records the objects, properties and relations that are salient at each point in the discourse in a linear stack (i.e., the top of the stack contains the elements in focus at any given point in the dialogue).

Although their model was designed to be able to address dialogue situations and dialogue phenomena like interruptions etc., in practice, there are some problems that arise even in monologue. For example, consider a story comparing two people, e.g. "Rose Red was a practical child while Snow White always had her head in the clouds. While Rose Red would pick berries for their mother, her sister would be sitting with her needlework forgotten in her lap"; the focus switches from Rose Red to Snow White at the end of the first sentence, but when it swings back to Rose Red in the beginning of the second sentence, the top of the attentional structure contains Snow White, and not Rose Red. Since one of the main motivations for the attentional structure is to aid in pronoun resolution, the algorithm would have trouble resolving "her" in the phrase

[5] Thanks to Johanna Moore for this point.

"her sister" in the second sentence to Rose Red and not Snow White, since the latter person is in focus. Penstein-Rose' (1995) gets around this problem by introducing a multiple-headed stack to model attentional structure.

However some basic problems arise in modelling intentional structure itself, both in monologue and dialogue, which are illustrated by examples in which speakers make digressions. Also, since Grosz and Sidner model intentional structure as a tree-structured hierarchy, a lower-level element can never serve more than one purpose in the discourse, e.g., it cannot be seen to be related to another parent, whether in the same relation or in a different relation. The very lack of any sort of account of rhetorical or semantic relationships means that the resulting description of the discourse is much less fine-grained. That is, while it accounts for which points might be subordinate or must linearly precede others in the text, it does not account for *how* the elements are actually related, and as we have discussed earlier, such an account is necessary in order to judge a discourse as coherent. Their coarsely defined intentional relations also leave much more determination of which relation holds up to the analyst; clearly a better solution would be to have more specific and fine-grained relations that leave less ambiguity in the assignment procedure of the relations.

In terms of dialogue, although Grosz and Sidner (1986) present an application of the theory to a task-oriented dialogue that accounts for different CPs' intentions, they formulate these intentions as highly dependent on each other, and assume (1) that something is a shared belief such that unless otherwise stated, (2) one CP will adopt the intention to perform an action the other CP intended him to, and (3) that in adopting this intention to carry out some action, the CP also intends to perform the necessary subactions. Grosz and Sidner note that in order to furnish a complete account of the intentional structure of a dialogue, one must be able to say "how the satisfaction of one agent's intentions can contribute to satisfying the intentions of another agent". However, while this might work in simple, highly cooperative task-based dialogue with minimal mixed initiative (i.e., a task-giver and a task-follower), it seems unlikely that it will work in non-task-based dialogues, and especially ones in which the CPs are in conflict or have conflicting views; while their intentions might be cooperative at the level of attempting to maintain a coherent dialogue, they might really vary in terms

of how they relate to each other at more fine-grained levels. Their strict dominance and satisfaction-precedence relations also will not allow CPs to change their minds or negotiate, and since dialogue is fundamentally a dynamic phenomenon, this is very problematic. Grosz and Kraus (1999) address some of these issues by modelling collaboration via SharedPlans.

2.3.3 The PTT Model of Dialogue

While the Information-State (IS) framework for dialogue that the PTT model is based upon allows freedom to model social commitments (SC, or SCP, where P stands for propositions), dialogue history (DH) and common ground (CG) in its attribute-value matrix (AVM) structured representation, the PTT model (as described in Matheson *et al.* (2000)) does not aim to cover the range of dialogue that can be represented by the IS model. For example, disagreements cannot be modelled, since CPs only have minimal models of the other CP's IS, and updates do not currently involve checking the model of the other CP's IS for misalignments. Furthermore, the taxonomy of conversational acts (CAs) described in Poesio and Traum (1998), and in particular, the core speech acts (SAs) capture predominantly illocutionary or intentional actions, and make no mention of the semantic relationships between utterances. This means that the argumentation acts (AAs) that are built from the SAs do not take into account semantic relations at all. While PTT supports an INT field representing grounded intentions (contained inside the GND field), these intentions can only be related via immediate dominance and unary predicates labelling them as satisfied or dropped (see Kreutel and Matheson (2001b)). This means that the only ordering of intentions possible in PTT is as a set of two-level trees, (not too different from the Grosz and Sidner representation). However, semantic relations relating propositions in the DH need to be accounted for (possibly more than intentions themselves) when making context-dependent updates among other things. For example, Kreutel and Matheson simply list propositions in the SCP field with no relations between them. This is fairly significant, since update scenarios and determination of argumentation acts cannot be context-independent, and scenario alone might not provide enough context.

Another subtlety is that the agent cannot question preceding premises; it may be

possible to formulate the questioning process as trying to achieve an epistemic goal (as Knott formulates it, which is discussed in the section below), but to do so, you have to have a clear goal (or possible counter-example in mind), and the questioning process only involves posing the goal, not then going on to try and determine its truth. Furthermore, the goal/counter-example itself may not be known by the agent when questioning a preceding premise/proposition–the agent may simply doubt its plausibility given other things she knows about the world.

2.3.4 Knott's Approach to "But"

Knott proposes an algorithm that simulates the mental state of an agent recognising contrast signalled by "but" in Knott (1999a). He presents algorithms for two relations signalled by "but", i.e., the *defeated-expectation* and *failed-plan* senses of "but". For Knott's *defeated-expectation* (or *epistemic*) interpretation of "but" (Knott, 1999a), he assumes the agent has a goal to answer the question whose answer is given by the "but" clause. So, e.g., for "Bob was up with the baby all night, but he looked as fresh as a daisy", the agent has the goal to determine whether Bob is tired. He then needs to backward-chain on the proposition "Bob is tired" to find a defeasible rule whose left-hand-side (LHS) is a known fact, e.g., "Bob was up all night". The agent then concludes defeasibly that Bob is tired; he then decides to ask whether Bob is tired perceptually as well, and then finds that Bob does not look tired, cancelling the defeasible expectation that he should be tired. This is an example application of Knott's expectation-based "but".

In Knott's example, what one needs to perceive to cancel the defeasible expectation is present in the text (i.e., "he looked as fresh as a daisy"); one wonders what one would do if a perception necessary to cancel the expectation were not present in the text itself. Although Knott wants to draw attention to the lack of decomposition of traditional presuppositional accounts of "but" in this usage, it remains to be investigated how one would constrain inference in cases where explicit evidence is not provided in the text. However the idea of explicitly representing the semantics of a connective with an algorithm simulating an agent's mental state as he processes the sentence, while similar in some ways to the PTT account, is also quite a novel approach and will be

the basis for our approach to the various relations cued by "but" in dialogue. We will develop a similar approach in this thesis to account for how the arguments of contrast and/or the nature of the contrast are determined algorithmically, which fits in well with the idea of dialogue as a dynamic phenomenon.

There is also some ambiguity in how the agent decides to backward chain on the affirmative answer (i.e., "Bob is tired") to the question he has as his goal (i.e., "is Bob tired"); it is not so clear how this algorithm would handle a positive resulting state, e.g., "Bob went to sleep early but he still looked tired", since this would presumably have the same goal to answer the same question (i.e., "is Bob tired"), but has the opposite outcome, which does not seem quite right, because if he then finds the fact that "Bob went to sleep early" he will get the right outcome (i.e., defeated expectation), but it could be a bit of a stretch (the LHS can in general spawn a lot of possible defeasible outcomes of varying degrees of specificity and relevance, and what if he found another rule whose LHS led to the opposite conclusion?)–altogether, the algorithm's search would probably be too unconstrained and does not sound very feasible computationally.

Furthermore, for dialogue in the examples contained in the appendix, although we can characterise utterances as being either demonstrated via *substantive* or *epistemic* planning algorithms, in some cases, actually running these algorithms might be quite difficult. Also many of the possible alternative situations and counter-examples introduced in the cross-speaker sentence-initial uses of "but" below are only remotely connected to the contrasting earlier clause. It seems that posing epistemic or substantive goals and trying to show defeated expectations or plans does not have enough breadth to address the use of cross-speaker "but" in spontaneous, non-task-related dialogue, in which "but" frequently relates alternative situations that are almost definitely not the result of negating a given expectation. Perhaps the problem with this approach lies in the negation of the consequent/expectation defeasibly inferred from the LHS; negation seems too strong to account for not-necessarily mutually-exclusive alternatives.

Summary

Here we presented two theories of discourse and text structure, i.e., RST and Grosz and Sidner's model, and argued why they would not be directly applicable to cross-

turn "but" phenomena, particularly in non-task-based dialogue. We then considered the PTT IS update dialogue model and addressed its shortcomings with cross-turn relations. Lastly we considered Knott's algorithmic treatment of "but". The main points made were:

- Necessary modifications to RST:
  - For dialogue we need to address multiple speakers' intentions, their co-operative interactions, and how they guide rhetorical structuring of their dialogue. This requires modifying the RST definition of relations, since this is currently defined as being guided by a single speaker's intentions. Presentational relations especially need to be modified for dialogue.
  - Informational and intentional relations need to hold simultaneously.
- Necessary modifications to Grosz and Sidner's theory:
  - There is no account of discourse relations between utterances, and intentional description alone cannot account for informational relations, i.e., there is no account of *how* the utterances are actually related, which is involved in determining local coherence.
  - Modelling the attentional structure as a single stack will likely encounter problems very quickly in dialogue situations, e.g., the Snow White, Rose Red example.
  - Their model will need extensive modification in order to address spontaneous conversational dialogue as opposed to task-based dialogue, since we need to be able to account for the sorts of phenomena that occur with mixed-initiative dialogue in which CPs have possibly conflicting views and less clear goals.
- Problems with PTT:
  - There is no mention of discourse relations between utterances here either, and determination of argumentation acts cannot be context-independent.

  - We need a fuller representation of the other CP's IS in order to model disagreements, different perspectives, etc.

- Problems with Knott's approach:

  - It is unclear how one would find appropriate goals with only text input.
  - It is also unclear how one would constrain inferencing.
  - Negation might be too strong; we probably need to be able to account for not-necessarily-mutually-exclusive alternatives, which would probably make inference even more unconstrained.
  - It is not clear whether all the different senses of "but" can be represented in just the violated-expectation and defeated-plan senses of "but". What about concession, for example, or disagreements?
  - Extending this to account for dialogue phenomena and contrast across speakers might not be so straightforward.

2.4 Some Necessary Additions to Dialogue Models

In the previous section we saw how existing models of discourse will not be able to account (in their current formulation) for cross-turn "but"-cued relations. Keeping in mind the underlying issue of assigning discourse relations where appropriate to establish how turns are related and therefore coherent, we now turn to some necessary additions to current discourse and dialogue models and address how these additional considerations enable modelling cross-turn relations. We start by focusing on grounding issues, and then go on to address necessary additions to semantic update processes. We then address bridging issues and discourse relations before finishing this review by considering the difference between opinion and belief and topic continuation.

Grounding

Grounding is the process by which DPs establish that they have mutual knowledge about the state of the conversation, in the roughest sense. It describes the process by which DPs establish (on-line, i.e. dynamically) that their utterances have been

communicated and understood in the course of the dialogue. Currently in the PTT model, DA information gets moved to GND only via explicit grounding acts which are achieved by *backward-looking* Understanding Acts in the DRI scheme, e.g., *signal-nonunderstanding* and *signal-understanding* via *repeat-rephrase*, *completion*, *acknowledge* and *correct-misspeaking*; no provision is made for implicit acknowledgement[6], which can usually be assumed to hold if the hearer goes on to continue the dialogue coherently without questioning or explicitly acknowledging the previous utterance.

The rationale behind explicit acknowledgement as the only form of grounding in Matheson *et al.* (2000) is that classical SA theory is inadequate in that it does not account for cooperative interaction in dialogue, and assumes that simply asserting something makes it mutually known (or grounded), a position which is not accepted by Matheson *et al.* (2000), who include explicit acknowledgement as a social commitment of DPs.

Implicit acknowledgement happens quite frequently in cooperative dialogue (probably more frequently in non-task-oriented dialogues than in task-oriented ones), and it needs to be accounted for in a theory of natural, cooperative dialogue. One approach toward incorporating implicit acknowledgement strategies into the view of dialogue as cooperative interaction would analyse the subsequent dialogue to determine if previous material has been implicitly grounded. Cautious treatment of implicit acknowledgement should distinguish between (1) continuing the dialogue without making use of the semantic content of the previous utterance and (2) continuations which do involve the semantic content of the previous utterance in some way. In the case where the previous utterance's semantic content is involved, one can probably determine if it has been understood. It is clear that understanding the previous utterance and demonstrating this understanding involves some form of grounding.

[6]In the Poesio and Traum model of dialogue, implicit acknowledgement needs an overt grounding act, while in the model by Cooper and Larsson, implicit acknowledgement is assumed; Cooper and Larsson's model follows an optimistic strategy of grounding, unlike the more cautious Poesio and Traum model at the other extreme. In Kreutel and Matheson (2001b), Kreutel and Matheson discuss inference rules for addressing implicit acceptance of answers in a question-answer situation, and this is the sort of approach we will follow.

2.4.1 Semantic Update

One issue has to do with how new utterances get merged with what has been said already. Poesio and Traum (1997) and Poesio (1995) show that evidence from spontaneous speech supports on-line processing of utterances, since turn-taking and grounding acts occur in between utterance fragments, indicating that the Common Ground is updated before a core SA is completed. Clark (1992) also argues that utterances, especially longer, more complex ones, are often grounded after each package of information is communicated. Continuation of discussion about the topic under discussion is handled in Poesio and Traum (1998) with a *continue* grounding act that simply concatenates the contents of the new DU to the evolving IS. Essentially continuation can be described as concatenation, since new DUs get put into the CDU, shifting back all the preceding DUs into the preceding fields (i.e., the old contents of the CDU go into PDU etc.).

Although in Traum and Hinkelman (1992) and Poesio and Traum (1997), higher-level argumentation acts are somewhat similar to rhetorical relations (as in Mann and Thompson (1988)), with acts like *elaborate, summarise, clarify, question-and-answer, convince* and *find-plan*, Traum and Hinkelman claim that these can be built out of combinations of simple core SAs like *inform, yes-no-question, check, evaluate, suggest, request, accept* and *reject*. Poesio and Traum (1997) argue that argumentation acts (the PTT equivalent to rhetorical relations) are related via *dominance* and *satisfaction-precedence* relations between conversational events which are then grouped into larger events known as *conversational threads* (some of which are also referred to as *discourse segments*).

Moore and Paris (1994) have made similar arguments linking intentions and intentional structure described by dominance and satisfaction-precedence relations to informational aspects of rhetorical structure (from RST). Although the earlier papers on the PTT model make provisions for rhetorical structure, in Matheson *et al.* (2000) there is no attempt to represent how new information is related to old information either intentionally or semantically when new DUs get processed; the new DUs are simply concatenated onto the evolving IS. This means that new utterances that attempt to correct or modify earlier utterances do not actually show their relationship to the old DU

that is being changed when they are added to the DH, and we lose the sense of how the DUs interact with each other in the evolving DH. This is especially significant for contrastive "but", particularly when it has a corrective function, e.g.:

(2.16) A: I think their last album, "Bossanova" was their best.
B: But I think "Trompe le Monde" was actually their last major album–have you heard that one? I liked it much better than "Bossanova".

Preserving the monotonicity of updates is probably a good idea, since it preserves the immediate context in the IS for any given point in the evolving dialogue like a snapshot in time; however, in order to note that CPs are modifying the information in their common ground, we need to introduce semantic relations linking what is being corrected to the introduced correction and note that this information is being corrected. If we did not account for the correction that B is making when adding B's DU to the IS, we would end up with contradictory information in the IS. Capturing B's correction requires noting which utterance in the DH's content B is actually addressing (B could be correcting something several turns back in the DH rather than the immediately preceding DU). So for the example above, we would need to introduce a relation something along the lines of $correction(B, A, PDU, [last\ album = Trompe\ le\ Monde], [last\ album = Bossanova])$. We might additionally want to note which DU in the DH contains the old information that is being corrected, which is why PDU is the third argument in the relation above; by default the corrected information occurs in the CDU, and as the dialogue evolves, this relation will indicate how many states back the old information occurred in, so if we need to check later how much of a lapse occurred before material was corrected, we can.

To get around the lack of a "merge" operation that incorporates the semantics of the incoming utterance into the DH and updates information in the DH in most dialogue models, we must first characterise how new material in incoming utterances interacts with information already in the DH[7]. This is a requisite step towards characterising how contrast across speakers in dialogue works, as we saw briefly in the example above. Poesio and Traum (1997) approach this by incorporating Muskens' Com-

[7] In theory, new information could interact with material explicitly present in previous DUs as well as with information inferred from the IS, and this will need to be addressed in the discourse update algorithm.

positional Discourse Representation Theory (CDRT) into the IS model of dialogue. They argue that while DRT is excellent at handling dynamic update and accounting for many semantic aspects of context, it abstracts away from pragmatic information and cannot process fragments which occur frequently in dialogue, necessitating their shift to CDRT to model semantic update. They contrast truth-conditional sentence meaning (e.g., from DRT) with intentional utterance use (e.g., SA theories in AI), and note that pragmatic accessibility is tied to discourse structure (e.g., consider Grosz and Sidner's attentional structure). On the other hand, they also cite Clark and Marshall's essay "Definite Reference and Mutual Knowledge" (in Clark (1992)) which claims that whether referring expressions are felicitous depends only on shared information; Poesio and Traum argue that this means that how specific utterances are related to task structure must be part of the common ground, which must therefore include pragmatic information as well as truth-conditional information, and they advocate incorporating pragmatic information into CDRT.

Moore and Paris (1994) and Asher and Lascarides (1998a) make the further distinction between compositional semantics of utterances and beliefs and criticise the representation of semantic interpretation and belief revision in the same module; Lascarides and Asher (1999) argue that one problem with this assumption is that interpreters then have direct access to other speakers' cognitive states, which is certainly not the case in human-to-human dialogue. (Possibly the separation between CPs' ISs in the TRINDI model of dialogue gets around this to some extent.) Asher and Lascarides argue that Asher's definition of contrast relations[8] constrains anaphora resolution, distinguishing coherent from incoherent responses in dialogues involving anaphora by constraining accessibility for embedded elements (see Example 1 in Lascarides and Asher (1999)).

They argue that since rhetorical relations constrain anaphora resolution in dialogue, a theory of discourse structure in dialogue should represent rhetorical relations, which they implement in their theory of Segmented Discourse Representation Structures (SDRT). They also argue that both the CPs' underlying goals and the linguistic structure constraints are necessary to resolve anaphora (and arguably interpret dialogue reliably), and that these modules must be separate but able to communicate. In

[8] Asher defines Contrast as having a partial isomorphism between two hierarchical DRS-structures with at least one pair of nodes with opposite polarities, (Asher (1993)).

Asher and Lascarides (1998b) they describe semantic and pragmatic constraints on answers to questions, thereby giving an account for coherence in the domain of question-answering. For example, their model predicts that B2 and A2 are odd responses to A and B1 in the dialogue below:

(2.17)
A: How can I get to the treasure?
B1: It's at the secret valley.
B2: ?Mary's hair is black.
A1: But I don't know how to get there.
A2: ?But I know how to get there.

However, their approach might be too rigid to handle conversational dialogue, where even utterances like B2 might be more likely to be coherent than in more formal, less spontaneous or task-based dialogue, given the appropriate context. Rather than blocking responses regarded as incoherent, we will advocate grading responses if possible, given the context of the dialogue and the DH, and prefer ranking coherence on a multi-valued scale rather than on a binary one.

We will also need to consider the extent to which the PTT model can be extended to keep beliefs and cognitive aspects of the CPs separate from linguistic information; currently the implementation described in Poesio and Traum (1997) models the common ground as a root DRS. While they also address the characterisation of anaphoric information and consider that information used to infer intentions should be kept distinct in the CG, unlike Lascarides and Asher (who do not mention CG explicitly or how it is incorporated in their model), they incorporate SAs into the CG, removing this separation. They argue that they are modelling sentence meaning rather than use, and go on to show that pragmatic accessibility is inextricably connected to discourse intentions and structure, citing Grosz and Sidner (1986) and others and providing examples of anaphora resolution.

2.4.2 Bridging Inferences and Semantic Relations

Some other considerations that will need to be accounted for in a discourse update algorithm for dialogue have to do with tracking topic continuity. Asher and Lascarides (1998a) define bridging to be "an inference that two objects or events that are intro-

duced in a text are related in a particular way that is not explicitly stated, and yet the relation is an essential part of the context of the text in the sense that without this information, the lack of connection between the sentences would make the text incoherent".

These ideas are related to the notion of *implicature*, which Levinson (1983) describes as inferences that are intended to be communicated but which are not logically entailed by the utterance itself. He goes on to partition implicature into *conventional implicature* and *conversational implicature*, following Grice (1975). Conventional implicatures are not derivable from Grice's maxims of conversation that encode special cases of his Cooperative Principle of language use, which assumes that communication requires cooperation between speakers, and states that speakers need to make their contribution as required at the stage at which it is required for the communication purposes they intend. The maxims generate inferences that are part of the noncompositional semantics of utterances, and conversational implicatures according to Grice are calculable by observing these maxims or deliberately flouting them for reasons in accordance with the Cooperative Principle. Grice further partitions conversational implicatures into particularised and generalised implicatures, where the former require a specific context in order to hold, while the latter hold in all contexts; an example of a particularised implicature is "let's close the windows" if "it's raining" is uttered and rain appears to be blowing in through open windows, while "I like some dogs" has the generalised implicature that I don't like all dogs, (or by the maxim of quantity I would have said so). Grice discusses "but" as an example of conventional implicature, claiming that the semantics of "but" is identical to that of "and", and the additional sense of contrast conveyed by "but" arises through conventional implicature. He characterises conventional implicature as having the following properties:

1. *noncancellable*; i.e., they do not rely on defeasible assumptions about context and so cannot be cancelled by explicitly denying what is been implicated.

2. *detachable*; i.e., they are tied to particular linguistic expressions or forms, and the implicature is lost if the wording of the utterance is changed.

3. not calculable by context or pragmatic principles like the Cooperative Principle.

4. not universal to other languages.

Conversational implicatures on the other hand he argues to be defeasible, nondetachable, calculable (from literal meaning and the Cooperative Principle) and nonconventional (not part of the conventional meaning of any of the linguistic expressions involved).

Oberlander and Knott (1996) argue that "but" can be viewed as launching conversational rather than conventional implicatures by showing how "but" is both nondetachable and cancellable, throwing into question its classification as a case of conventional implicature. They show that "but" triggers defeasible conversational implicatures of violated expectation and defeated plans (also addressed in Knott (1999b)); presumably its role in concession and semantic opposition can also be seen to trigger defeasible conversational implicature, and possibly for disagreements and corrections as well, though this will be left to further research.

Oberlander and Knott use the taxonomy of cue phrases proposed in Knott (1996) to generate partially-ordered sets which can be analysed in terms of *scalar quantity implicature*. Since Knott's taxonomy assigns feature-value bundles to discourse cues, he argues that these bundles of feature-values correspond to discourse coherence relations, which can presumably also be ordered on a salient scale like the cues themselves if the cues are viewed as cases of scalar implicature. They hypothesise that speakers use more general cues in cases when they wish to avoid affirming more specific relations, or when they can rely on context to license implicature to the more specific relation the speaker intends to communicate.

2.4.3 Additional Issues

Opinion vs. Belief

We will need to determine a systematic set of criteria for distinguishing rejection of semantic content from disagreement. It appears that while this distinction will also depend on the pair of utterances contrasted, the form B's reply takes might also provide clues; so for example, if B uses verbs that express propositional attitudes, this might be an indication of disagreement rather than a rejection of semantic content. Non-factive verbs might also be clues that the speaker is giving an opinion rather than stating a fact that they believe to be true. However these clues certainly do not exhaustively describe

statements of opinion. There are also cases in which a speaker believes what she is saying to be true, but expresses it in a tentative way by couching it inside the complement of a non-factive, e.g., "I think that Bush is the President of the U.S. now"; this sort of behaviour might be correlated partially with speakers' context or relationships, and we will need to be careful to avoid making across-the-board generalisations with some of these claims.

How then to distinguish between statements of belief (propositions speakers believe to be true) and opinions, which cannot be tested for rigorous truth commitment? In this example, one can probably use the value terms like adjectives as clues that this is a value-judgement rather than a fact believed to be true. The reason why this distinction is important will arise in grounding issues; it is possibly the case that if a CP expresses an opinion, they generally expect understanding but not necessarily agreement, or possibly it is the case that their opinion is less refutable than a statement of belief would be. In any case, it would probably be sensible to distinguish between opinion and belief in the IS, even if they do not get grounded differently. It would be fairly straightforward to represent different speakers' perspectives (i.e., their beliefs) in the common ground; for example, using belief (*Bel*) and mutually-believes (*MB*) operators, if CPs A and B are aware of each other's perspective on proposition *p*, then if $Bel_A(p)$ and $Bel_B(not\ p)$, then while it is not the case that either $MB(p)$ or $MB(not\ p)$ hold, it is the case that both $MB(Bel_A(p))$ and $MB(Bel_B(not\ p))$ hold. It might be the case that different contexts might require different grounding behaviours, for example consider the differences between a formal debate situation or a tutorial and an informal conversation between friends.

Topic Continuation

Determining whether (and how) the topic is being continued often involves determining discourse relations between the current dialogue unit and at least one of the preceding dialogue units (or inferences arising from them) in the dialogue history. Initially determining what is being grounded and what is new information should be helpful. The new information somehow continues the old topic (which simultaneously grounds it minimally by showing that the old information has at least been partially understood); the new information then continues this old information along a particular di-

rection, and is related to the old information in some way (i.e., there is a discourse relation involved). Hobbs (1985) talks briefly about notions of topic, and we will need to consider both notions of topic and local coherence in these cases.

In disagreement/agreement cases, it needs to be noted that an expression of opinion is being made. In utterances for which the same span of text both grounds and introduces new info, we will need to account for the possibility of multiple relations holding simultaneously. However, we have already seen several approaches which allow for multiple relations to hold simultaneously, e.g., consider Moore and Pollack's approach which allows informational and intentional relations to hold simultaneously, and the DAMSL multi-layer scheme which allows FCFs and BCFs to hold for the same span of text.

Another issue that has been discussed involves linking the degree to which the topic is continued and grounding of old information. Given the connection between grounding and coherence, we can see that a binary inclusion of elements into or out of the shared common ground is inadequate in many instances. We have seen that it does not allow the distinction between agreement/acceptance and understanding.CPs seem to have a sense of confidence of grounding which governs their dialogues; well-grounded propositions can probably be assumed to hold and be built upon without being too cautious, while less well-grounded propositions might require more clarification and some cautious exploration to determine whether they are really understood. One domain in which one really sees explicit grounding strategies is in the BE&E[9] tutoring dialogues, where students hardly ever ground what they say and tutors almost always initiate grounding procedures to check that the student has really understood what the tutor is trying to communicate (Tsovaltsi (2001)).

One general problem with DRI tags is their narrow range; they only allow a limited number of tags that focus primarily on interactional games of DPs, e.g., FCFs pose opening moves, and BCFs hold completing/responding moves, and they cannot characterise many of the other aspects of relationships between utterances. For example, BCFs do not allow for continuation of topic under discussion, or the relation of what is being said to what went before. There is not even the possibility for annotating

[9] "Basic Electricity and Electronics Corpus"; see http://www.cogsci.ed.ac.uk/~jmoore/tutoring/BEE_corpus.html.

the sort of opinion of preceding material that is expressed. There needs to be some sort of way to formulate discourse relations in dialogue, both intentional and informational relations, in order to more fully characterise how utterances are related to previous ones, and in order to extract and distinguish between both informational and intentional views on the topic under discussion for both DPs.

As Moore and Pollack (1992) argue, if we are using RST relations, this is likely to involve multiple relations, since it is often possible to recognise different relations depending on whether one reasons from intentional coherence to informational coherence, or vice versa. In the case of contrast, consider a modification of their example, "George Bush supports big business. But he's sure to veto House Bill 1711". Here although the "but" indicates that there is some contrast between the two sentences, it is still possible to recognise alternative relations. For example, let us assume that the DPs are debating, and have opposing viewpoints on whether Bush will veto bills against big business. If the hearer knows that the speaker is arguing that Bush will veto more bills, then he might reason that S_1 is extra evidence for S_2, since the "but" cues some sort of contrast between vetoing House Bill 1711 and supporting big business. In fact if the speaker does not know what House Bill 1711 proposes, she might additionally reason that it is in opposition to big business, since there is a contrast between the two propositions; this is an example of reasoning from intentional to informational coherence. On the other hand, suppose that the speaker knows that House Bill 1711 opposes big business; then the hearer will deduce that the speaker intends to convey a contrast between the two propositions (assuming that the speaker assumes that the hearer knows what Bill 1711 proposes), reasoning from informational to intentional coherence.

Notice that both ways of recognising what relation is involved requires recognising what is in contrast and how. However in the first situation, one recognises both contrast and evidence as holding simultaneously, while in the second case one only recognises contrast as holding. So this just shows again that the uniqueness claim that RST makes is questionable. Probably the whole paradigm on which RST is founded will have to be changed in order to handle two speakers with separate intentions and goals, since RST was designed to hierarchically represent a single speaker's intentions in organising a text.

It is clearly necessary to be able to characterise the semantic and pragmatic relationships between utterances, even when they are between speakers, in order to fully characterise the interaction.

Summary:

In this chapter we introduced current relevant work on contrastive discourse relations, going on to present several seminal approaches to discourse and dialogue modelling, addressing where their shortcomings lie with respect to modelling cross-turn discourse relations, and finally presenting some additional considerations for dialogue models that account for cross-turn relations. We saw that all the approaches to modelling both discourse relations and dialogue fail to cover cross-turn "but"-cued relations or address coherence relations across turns in dialogue, particularly when these coherence relations need to account for inferred expectations or relations with respect to a TC. Furthermore, we learned that "but" often involves disagreement or minimally a difference in perspective, which requires an approach which addresses these conflicting beliefs and intentions. We found that in order to interpret cross-turn "but" relations we need to analyse multiple speakers' intentions, beliefs, plans and expectations. These problems indicate that we need to develop an approach for analysing cross-turn "but" which, while it rests on the foundations of dialogue and discourse theories, needs to account for the novel problems involving cross-turn coherence relations. Modelling cross-turn "but" will in turn lay the groundwork for modelling other cross-turn relations, which will contribute to establishing the coherence of how speakers respond and communicate in dialogue. In the next chapter we will look more closely at cross-turn "but" phenomena by investigating examples from dialogue corpora before we turn in Chapters 5 through 7 to modelling cross-turn "but" cued coherence relations themselves.

Chapter 3

Investigating the Data

This thesis aims to determine how turns in dialogue are coherent when there is a turn-initial "but". The idea is that this "but" signals a discourse relation between the two turns which needs to be interpreted in order to make the dialogue coherent. In this chapter we will turn our focus to examples of turn-initial "but" in dialogue corpora to base our study from real examples of "but" relations and learn some of the characteristics of these examples in order to determine characteristics of the underlying relations communicated.

This chapter describes our initial efforts at investigating "but"s in dialogue corpora in order to motivate hypotheses about what turn-initial "but" communicates and how the "but" turn relates to the previous turn. As stated in Chapter 1, we assume utterance-initial cases are most likely to involve cross-turn relations. We saw there that in the Switchboard (NTOD) corpus, "but" appears utterance-initially about 23.6% of the time when it appears (excluding cases where it follows phrases, false starts or single-word agreements or disagreements like "yeah, but"). In the investigation described below, we also considered cases which involved "yeah, but", "OK, but", etc., assuming that these cases also relate to the previous turn.

The corpora fell into two broad groups: Task-Oriented Dialogue (TOD) and Nontask-Oriented Dialogue (NTOD), and this chapter focuses on distinguishing the different characteristics of these two types of corpora where "but" is concerned. We distinguish TOD from NTOD on the basis of whether or not the dialogue participants are involved in planning or performing a task together. In particular, TOD involves task-plans which

the speakers pursue through their conversational actions. However, dialogue participants engaged in a task together can sometimes introduce new and unrelated topics or digress briefly into nontask-related conversation which should be classified as NTOD. Likewise, in NTOD, speakers can issue commands, involving mini task-plans and then pop out of the TOD back into their NTOD conversation. In order to be stricter about how we distinguish the two, we will consider the parts of an utterance turn relating to the task as TOD and those that do not address the task as NTOD.

In this chapter we will start by presenting a pilot study intended to provide an initial comparison between TOD and NTOD. We will then try and be more systematic about the effects of context by presenting a procedure for determining what is in context to provide some intuition about the different aspects of meaning involved in cross-turn "but" cases. We will then analyse examples from the TRAINS TOD and Switchboard NTOD corpora to gain some insights into how cross-turn "but" relates material in these cases. We will finish by drawing some conclusions about how TOD and NTOD are distinguished.

3.1 An Initial Comparison Between TOD and NTOD

In this section we will present a pilot study which will help us draw intial distinctions between TOD and NTOD. The goal of this comparison is to analyse a small portion of TOD and NTOD corpora in order to learn how cross-speaker "but" differs in each (both in terms of how things are related and what is related). We look at examples of cross-speaker "but" in the TRAINS[1] and Monroe[2] TOD corpora (although the Monroe corpus was considered to a lesser extent than TRAINS), and the Switchboard[3] corpus. These corpora were chosen because they all feature cases of turn-initial "but" in dialogue and were readily available. Furthermore, they span TOD and NTOD and therefore give a range of different types of dialogues. As mentioned earlier, we identify cross-turn "but" as those cases involving turn-initial "but", where by turn-initial

[1] See www.cs.rochester.edu/research/trains for more information.

[2] See www.cs.rochester.edu/research/cisd/resources/monroe/ for more information.

[3] Switchboard NTO telephone-based dialogues on randomly assigned topics between strangers. See www.ldc.upenn.edu/cgi-bin/lol/swb for more information and the corpora themselves.

we include turns in which the “but” is preceded by a word, phrase or false start, so this includes “yeah, but”, “um, but”, “the flight, oh but” etc. The aim is to determine what is related (and how) by cross-speaker “but” in the different TOD and NTOD corpora.

There has been a lot of work in monologue on ”but”, but not in dialogue. Since comparing its appearance in monologue and dialogue could form the larger part of a thesis itself, we will eschew looking at monologue “but” in the corpora and so will not attempt to distinguish monologue “but” from its dialogue usage. We will focus our attention instead on dialogue only in the hope that a broad typology of dialogue-specific uses of ”but” can emerge.

Some initial speculations

A first question one might ask is whether “but” is more common in TOD or NTOD. One possibility[4] is that TOD might have more ”but”s, since it is often more direct and less polite than NTOD. On the other hand, ”but” might be less formal, and so more common in NTOD[5]. Statistically ”but” is the 20th most frequent word in the Switchboard NTOD corpus, which has a vocabulary size of 41077 distinct words, and it appears utterance-initially 23% of the time when it appears, even excluding cases when it follows a single word. From the Monroe task-oriented corpus, Stent reports more than 50 utterance-initial ”but”s (0.09% of the corpus) with 232 occurrences total in the corpus (so there are around 0.45% “but”s). About 25% of the time it is utterance initial, which is comparable to Switchboard, but it is only the 58th most frequent word, of a much smaller vocabulary size (1550 distinct words, of a corpus of a total of 52000 words). Normalising by dividing rank by vocabulary size to make sense of the numbers spreads them even further apart, and it appears as if ”but” is much more frequent in Switchboard than in Monroe. This does not make the prospect of studying TOD for turn-initial “but” seem promising at a first glance, as Monroe is a TOD corpus specifically designed and collected to include more rhetorical relations than in most other TOD corpora.

However Knott (1999b) indicates that turn-initial “but” in TOD can be analogous to its defeated expectation occurrence in NTOD, since it is possible to link the failed precondition of a plan in TOD to failed defeasible expectations entailed in NTOD. It

[4]Thanks to David Traum for this idea.

[5]This was an idea of Amanda Stent’s via email.

might also be the case that other TOD corpora might have more ”but”s because of conflicting agent goals and plans. From an initial study of both Monroe and TRAINS dialogues, some representation of plans will be necessary to annotate what is in contrast in these dialogues, because the ”but” often contrasts with the plan that the agents are building.

3.1.1 An Initial Annotation Attempt

Since our goal is to determine what information about the contrasted speaker turns allows us to distinguish between different relations (e.g., DofE and concession), we first consider previous annotation schemes for TODs to determine whether these schemes contain information that will be relevant for distinguishing between these relations. The DAMSL scheme (Allen and Core, 1996) annotates dialogue acts (DAs) in several layers, accounting for (among other things) forward and backward looking functions that a given utterance unit may perform and describes how one distinguishes between these functions. For example, this scheme allows one to annotate an utterance unit (within a speaker turn) as both an assertion or information-request (forward looking functions) and also an agreement with partial acceptance. Sikorski and Allen (1997) propose an additional scheme to mark higher-level problem solving actions in TOD. The Penn Discourse Treebank[6] project presents an annotation tool which enables annotation of argument structure for (among other things) connectives. Their goal in annotating the argument structure of connectives is to support the extraction of inferences associated with particular connectives. However while their tool results in the determination of the connective’s arguments, their focus is more on syntactic features of arguments and differences between syntactic and discourse structure congruences as a result of attribution (Dinesh *et al.*, 2005). Also they analyse the Wall Street Journal corpus of news articles, which does not involve the fragments, false-starts, etc. of dialogue or the differences in perspective between turns which we focus on when addressing cross-turn coherence relations.

However, none of these schemes annotates utterance units with lower-level plan-

[6] See *www.cis.upenn.edu/~pdtb* for the D-XTAG annotation scheme, papers, and other information on the project.

ning operations to note whether they describe preconditions, effects, actions, or goals in the joint task-plan being formed in the dialogue. We hypothesise that this information about what planning operations are being communicated might help distinguish which relation/s to generate in a given situation.

To this end, we performed an initial pilot study which involved annotating 20 examples each of both the MAPTASK and TRAINS TODs.[7] We searched for cross-speaker "but" examples which contrasted with material in the preceding speaker turn (ignoring turns involving back-channelling and utterances that signalled understanding). We then annotated both the relevant preceding utterance unit and the utterance unit containing the "but" with DAMSL tags, problem solving actions (Sikorski and Allen, 1997), and the planning operators involved. The annotation involved (for each utterance unit) noting (1) the mood of the unit (i.e., is it a command, question, or assertion; while this does not involve planning operators, it was annotated because it reflects syntactic considerations which do not appear in either of the other two annotation schemes used), (2) whether the unit describes an action, effect, goal, precondition or constraint in the evolving joint-plan, and (3) whether this planning operation occurs in the present, past or future.[8] In Section 3.2.1 we describe the annotation scheme in more detail and in 3.2.2.1 we present some annotated examples.

Of the 20 examples in each corpus, Table 3.1 shows the distribution of cases which were omitted or were not easily classified, clarifications or corrections, cases of DofE or concession or both.[9] The examples were chosen only if they clearly involved cross-speaker contrast (i.e., they did not relate material within the turn)[10] and were classified by the author.

The corpora differ interestingly in the DofE:Concession ratio, with many more cases of DofE than concession in MAPTASK and vice versa for TRAINS, possibly because of the difference in corpora; TRAINS involves agents determining the optimal plan for actions that will subsequently be carried out while MAPTASK involves agents

[7] MAPTASK dialogues can be browsed interactively at www.ltg.ed.ac.uk/ãmyi/maptask/demo.html.

[8] Since this was a pilot study only intended to give some rough indication of trends, only the author marked the two dialogues so no measures of inter-coder agreement are available.

[9] See Chapter 4 for definitions of the labelled relations.

[10] To determine whether material within the turn is not related, the turn on its own should not be coherent (in terms of the "but"); i.e., the turn should only be coherent given the preceding turn.

Table 3.1: Relation classification of examples with count;percentage

Corpus	Omitted	Clarification	Concession	DofE	Concession & DofE
MAPTASK	7;35%	3;15%	0	9;45%	1;5%
TRAINS	2;10%	2;10%	12;60%	1;5%	3;15%

interleaving their speech and task-related actions rather than planning future ones. The fact that at least half these cross-speaker relations are either cases of concession or DofE is a good motivation for modelling these relations, though a study involving randomised selection of cross-speaker examples in multiple corpora, both TOD and NTOD and annotated by more than one annotator will be necessary to establish this indication of these relations' frequency in cross-speaker "but" cases.

3.1.1.1 Annotating DAMSL Tags

We then examined the annotated examples of concession and DofE to see if there are any annotation trends we might infer from DAMSL tags which might help us to predict which relation to interpret, shown in Table 3.2, where "T1" refers to the preceding speaker turn and " T2" to the turn containing "but" which responds to T1. The tags in the table are abbreviated from the DAMSL scheme, so Commit commits the speaker to some future action, while AD (action directive) and OO (open option) influence the listener's future actions. Reject and Accept include partial rejection and acceptance, IR is an information request and Answer answers a prior question. Entries in Table 3.2 correspond to DofE;Concession values, and the first two rows contain TRAINS counts for the tags while the latter two rows contain MAPTASK counts. For now we simply consider individual tags rather than combinations within or across turns; this is left for future work when more examples have been annotated. Tags with only one or two occurrences have been omitted from the table, and only counts for cases of DofE and concession are shown. Recalling that TRAINS had a higher incidence of concession and MAPTASK a higher occurrence of DofE, the two corpora ad also the incidence of DofE compared to concession for a given tag should not be compared. Rather we should compare the relative frequency of tags for a given relation (i.e., DofE or con-

Table 3.2: DAMSL annotations: Counts correspond to DofE;Concession classification

Corpus & Turn	Assert	Commit	AD	OO	Reject	Accept	IR	Answer
TRAINS T1	4;9	2;1	0;2	0;2				
TRAINS T2	4;15			0;1	0;4	1;8	0;1	0;1
MAPTASK T1	2;1		1;0				7;0	
MAPTASK T2	4;1	1;0	3;0			0;1	1;0	7;0

cession) and corpus. Recall also that since the occurrence of concession in TRAINS is much higher than its DofE incidence, and vice versa for MAPTASK, we should consider the DofE counts from MAPTASK and the concession counts from TRAINS as more indicative of any trends than the other two sets of counts. Concession in TRAINS usually involves assertions in both turns (9 cases) as indicated in Table 3.2, which also seems to be the case for DofE (with 4 cases of assertions in both turns). In MAPTASK it appears as if DofE tends to involve information requests followed by answers (although assertions were also quite common in T2). The counts for concession in MAPTASK were so low that we will not take them to be indicative of anything. So the overall trend is that concession tends to involve assertions in both turns, or assertions followed by rejections, while DofE tends towards information requests followed by answers.

3.1.1.2 Annotating Planning Operators

The next question to ask is whether our hypothesis that more planning-operator-specific information can help predict relations is valid. We count the number of operator pairs for each corpus and relation ("Conc." stands for concession) in Table 3.3[11] . Since we only consider 20 examples in each corpus, not all of which are classified as either DofE or concession, the data is sparse. The reason only 20 examples were annotated for this pilot study is because they were the only ones found in these two corpora (i.e., there were a total of 20 examples of cross-turn "but" in TRAINS and MAPTASK combined). For that reason, operator-pair counts for TRAINS DofE and MAPTASK

[11]The abbreviations in Table 3.3 are as follows: A for action, E for effect, G for goal and P for precondition or constraint.

Table 3.3: Counts for planning operator sequences (Turn 1 operator, Turn2 operator)

Corpus & Rel.	A,E	A,P	G,G	E,P	E,E	A,A	P,P	P,G	P,A
TR. Conc.	1	4	1	4	1	1	1	1	1
TR. DofE		1		2			1		
MAP Conc.							1		
MAP DofE		1		1		3	3		1

concession should be ignored since there were so few examples.

However for the more numerous corpus-relation pairs, it appears as if concession (from TRAINS counts) seems to involve goal–goal or effect–effect sequences more frequently, and DofE (from MAPTASK counts) seems to involve precondition–goal or precondition–precondition pairs. While these inferences are highly speculative since this is only a pilot study, they seem somewhat reasonable. For example, in concession, speakers are offering alternatives with respect to a salient topic or question under discussion, and it makes sense that they should offer alternatives of the same type of planning operator (e.g., alternative goals). Similarly, since DofE involves denying expectations, it makes sense that preconditions give rise to expectations, and what is denied is the goal's achievement or the validity of another precondition given the one in T1.

3.2 Being Systematic About the Effects of Context

Another issue to keep in mind when analysing examples from corpora is to be systematic about how much prior context is used in determining what is in contrast. In this section we will present an informal annotation scheme which helps distinguish what is in contrast in these cross-turn "but" cases.

In terms of annotating the contrast itself, we have used double brackets around the phrase or clause that is one of the arguments of contrast, e.g., "[[argument 1]] BUT [[argument 2]]". However, in many cases there might be additional or intervening clauses that are satellites (using RST terminology, Mann and Thompson (1988)) of the nuclear clause in contrast; in these cases we include these subsidiary clauses in the

double bracketing to indicate that it is the whole structure in contrast. In some cases the intervening or subsidiary clauses might not add anything at all to the coherence of the contrast relation, and in these cases, we leave them out of the double bracketing. We have been very vague here about how one determines whether the subsidiary clauses have any noncompositional effects on the contrast relation. According to compositional hierarchies of discourse relations like that proposed in RST, these relations would per force be included in the bracketing. We leave them out if they can safely be skipped without changing the meaning of the contrast at all in order to get at the essence of what is involved in determining the contrast relation without the additional clutter introduced by these subsidiary relations and clauses.

Our goal in developing the simple annotation procedure described below is to gain insights into the elements of the turns being related and thereby learn how contrast works in cross-turn "but" cases. The role of this analysis is simply to help the interpreter to gain some intuitions about the different aspects involved in different examples in order to guide the analysis. That is, since we have not tested intercoder agreement on the procedure below, we only present it as an informal tool to aid the interpreter and not as a formal annotation procedure. Testing agreement and establishing the reproducibility of the procedure to establish it as a formal annotation methodology are left for future work.

Turning now to our goal of isolating and analysing which elements are in contrast, we present some initial considerations which have been inspired by more lexically focused annotation attempts on "other" anaphora (Nygren (2004) and Modjeska *et al.* (2003)). While the phenomena being examined here is lexically triggered, the range of possible contrasts it can coherently license is vast compared to the restricted domains of "other" anaphora and NP metonymy in monologue, so while we can use their basic rationale for our analysis methodology, we need to formulate our own approach for the analysis of "but" contrasts in dialogue. We start by assuming a structural pattern that roughly corresponds to *A: X. B: But Y.* Any text before "but" in B's turn will be considered to be part of Y, so in these cases, Y looks like "Y1 BUT Y2". Given this rough segmentation, we will focus the analysis on the following elements:

1. The features of X
2. The features of Y

3. The relation between X and Y, whether this relation holds on the semantic or pragmatic level, and whether it takes generalisations of X and Y as arguments as opposed to the discourse segments containing X and Y; we are being vague about what generalisations consist of here since this might involve bridging inferences, etc.

When we focus on the features of X and Y, we will want to consider lexical triggers and similarities, syntactic patterns, semantic intuitions on the types of things in contrast and how they relate to each other, rhetorical relations between X and Y, and possibly Speech Act and Illocutionary Force effects. We will distinguish propositions, discourse segments and topics of discourse segments, note whether elements are evoked situationally or contextually, whether it is inferred, and if so, how it applies to the speaker's beliefs, and we will distinguish assertional and inferred elements of the utterance.

A very important issue to consider when analysing these examples is how information (of the sorts listed above) is raised in X and Y and how they relate to information in each other. We can then analyse these aspects of the dialogue to determine what inferences can be made about expectations and beliefs of the participants (and how they were interpreted) and more generally about what is being communicated.

Implicatures

One type of inference is implicature. Grice (1975) argues that "but" conveys conventional implicature, since in addition to the truth-conditional semantics of conjunction it shares with "and", "but" also introduces a sense of contrast between the conjuncts. According to Grice, implicatures are aspects of meaning that contrast with what is said or expressed truth-conditionally, and while conversational implicatures are calculated on the basis of Grice's maxims and his principle of cooperativity, conventional implicatures arise lexically rather than from the maxims.

Clearly we will want to extend our classification of inferences to categorise other sorts of inferences and implicatures also. The more clearly outlined our classification of inferences is, the more we will learn from this annotation exercise, since as we have learned from Grice, the contrastive sense of "but"'s semantics arises nontruth-conditionally and must be inferred. At the level of semantic relations between utterances and fragments, we will adopt RST's subject-matter relations but will adapt them to account for dialogue phenomena not accounted for by RST. For example, consider

the following dialogue:

(3.1) A: Helen didn't come to the party.
B: But I'm sure I saw her there.

Discourse Relations

Applying RST relations like contrast, concession or antithesis to cross-speaker utterances seems to run into problems intentionally, since the satellite and nucleus (or two nucleii, in the multinuclear case) are no longer spoken by the same speaker for a single intention or communication goal. In Moore and Pollack (1992), the distinction is made between informational and intentional relations. One way to get around the cross-turn problem involves labelling informational relations across speaker turns only, rather than intentional ones, since the former relate content only. Presumably to interpret (and annotate) at the intentional level, a model of SharedPlans, Joint Activities or agent communication languages would be more appropriate than RST. However even given this distinction between informational and intentional relations, the informational relations might still be affected by the dialogue context, and it might not be the case that dialogue information is strictly compositionally additive to informational relations; i.e., it might be the case that the dialogue context affects the semantic interpretation between the utterances.

Bearing these considerations in mind, we now present an informal procedure to aid humans analysing the source of contrast given a "but" turn. The goal of this exercise is to try and identify the various features of examples encountered in the corpora in order to learn more about "but" triggered contrast.

1. Determine the features of X and Y separately (and jointly, as appropriate) according to the following guidelines, where X represents the first turn and Y the second turn (i.e., Y contains the "but"):

   (a) Represent material explicitly present in first-order predicate logic form, classifying cues, connectives and properties (e.g., "duration") as predicates taking arguments.

   (b) Determine which aspects are inferential and classify them according to how they are triggered and what is entailed when possible:

- if they give rise to implicatures, classify them as conventional or conversational and give triggering constructions or Gricean maxims involved if available
- if they otherwise add to the non-truth-conditional semantics of the utterance, state how so (e.g., presupposition) and what added meaning arises

(c) Apply the following guidelines to all categories separately unless stated otherwise

i. Do X and Y relate assertions, questions or commands?

ii. Determine which propositional aspects of X and Y respectively are *old* and which are *new* pieces of information based on the dialogue history.

iii. Determine Illocutionary Force (IF) and/or Speech Acts (SA) performed in terms of how X and Y are related

iv. Determine *communicative intentions* of X and Y. This will often take the form of intentions to commit a given SA[12]
Intentional Structure: Are there any obvious specifi c satisfaction-precedence or dominance relations (Grosz and Sidner, 1986)?

v. Planning History: What salient and signifi cant goals, preconditions or effects were communicated in the dialogue so far? Label all planning information with type, i.e.:

- *Goals* take actions as arguments
- *Preconditions* are subgoals and so index their immediately dominant goal as their fi rst argument.
- *Effects* take the goals they are a result of as their fi rst argument.

vi. Planning Information: What goals, preconditions or effects are communicated in X and Y specifi cally?

vii. What is/are the main Topic(s) Under Discussion?
Determine what RST subject-matter rhetorical relations hold between SAs. Prefer relations between adjacent planning operations/SAs in the dialogue.

viii. Determine what salient aspects of the paraphrased dialogue help determine what is in contrast. Look for lexical or syntactic clues that were used to

[12]Following Grosz and Sidner (1986) and arguments by Kreutel and Matheson (2001a) that SAs are bids or offers until grounded.

determine the discourse relations involved.[13]

2. Summarise findings to explain how X and Y are related at the different levels of granularity.

We will use these guidelines to analyse a few examples from the TRAINS and Switchboard corpora, some of which were annotated earlier according to the simple scheme we used in the pilot study in Section 3.1.1. These guidelines enable a more in-depth analysis of the examples than we got with the simple notation of sentence mood, DAMSL tags and planning operators which we used in the pilot study. We hope that this analysis will shed further light onto the nature of "but"-contrast and possibly also reveal some low-level clues that co-occur in the various cases and can be used to distinguish between various kinds of contrast. Subsequent analysis would require an evolving representation of the plan the agents are forming, along with information about who proposed which parts of the plan in order to formally represent what is in contrast.

3.2.1 The TRAINS corpus

In this subsection and the next we will analyse examples from the TRAINS TOD and Switchboard NTOD corpora to examine more closely how "but" relates material across turns in these two domains.

The TRAINS TOD corpus involves agents planning how to move cargo between destinations efficiently, and involves planning efficient routes to achieve their task. There are 97 cases of "but" in the TRAINS corpus (this includes both turn-initial and non-turn-initial cases), which, given that the corpus contains 58,298 words in all, is a frequency of 0.16% (which is not insignificant, given that the frequency of "the" in the corpus is around 2.57%). The TRAINS corpus contains 859 distinct words in 98 dialogues and 6163 turns[14].

From an initial look at TRAINS, contrast occurs a lot with the evolving plan and gets signalled by "but", both turn-initially and non-turn-initially. As mentioned earlier,

[13] I.e., are there any lexical triggers or patterns or syntactic trends that seem to correlate with the contrast? These can be evaluated for their effect on the contrast via substitution tests.

[14] From Peter Heeman's thesis, p.62, Heeman (2004).

annotating this contrast requires some representation of the evolving plan which is accessible for both speakers to comment on/analyse/contribute to. Since most ”but” contrasts in TOD and TRAINS in particular seem to involve contrast with meta-level constructs rather than something explicitly present in speech, it might be more difficult to determine some general principles of analysis of these constructs that apply to more shallow approaches as well. Also, since this sort of contrast is with meta-level evolving constructs, it is much more of a global contrast than a local adjacency-pair contrast between adjacent utterances.

3.2.1.1 Some annotated examples

Here is an example of concession (TRAINS d93-13.1):

```
utt24 : u:  [two hours so that's <sil> seven hours in all] <sil> + with +]
utt25 : s:  + [yes] + <sil> BUT [it takes an hour to load the oranges]
```

Paraphrased d93-13.1:

```
utt24 : u:  [two hours]
            [so that's seven hours in all] with
utt25 : s:  [[yes]] BUT [[it takes an hour to load the oranges]]
```

Here ”yes” concedes u’s point, but BUT introduces an objection to u’s point. Although one could argue that this sort of case (of which we will see many more in the thesis, particularly in Chapter 6) involves an intra-turn relation, we model it as cross-speaker concession, because one of the relata of the concession turn (i.e., “yes”) depends crucially on interpreting the prior turn for its meaning. Certainly the “yes” does not significantly change what is being communicated by the “but” turn. “OK” and “yes” are frequently also used as acknowledgements, so they may not explicitly indicate agreement (this is more the case for “OK” than “yes”). Interestingly “No but” cases appear to be very rare if acceptable at all; in all the cross-turn “but” examples from the corpora considered in the course of this thesis (probably around 50 altogether as a lower estimate) none contained “no but” whereas “yes but” and “OK but” were common. “Yes but” and “OK but” might also function as politeness indicators, as they

soften the bluntness of a bare “but”-fronted turn. We did not find enough examples to indicate more than the roughest of trends, but from the analysis of all 20 of the examples in the pilot study, all involved material in front of the “but”, which ranged from “eleven (backchannel/thinking aloud), but” to “uh but”. While the “yes” and “OK” could be seen as optional, we will not argue that it is either optional or mandatory here; certainly in some cases it seems to simply acknowledge the prior turn, while in others it seems to soften the criticism following “but” by indicating that the speaker is in favour of the prior turn’s proposition but wishes to add additional information. We consider both cases to involve cross-turn “but” since even in the case involving explicit agreement, the prior turn must be interpreted in order to determine what is in contrast on a literal level, and it is on this literal level that the “but” speaker contrasts the previous turn with her turn. Even though on a meta-level the contrast may lie within the turn between agreement and criticism, there is no contrast between the literal “yes” and the rest of the “but” turn; even if the speaker had said “I agree but ...” we would still need to interpret the previous turn to see what was agreed with. That is, the “yes but...” turn itself cannot stand in isolation and be interpreted.

In any case, concession seems fairly frequent in TRAINS, and notice that here the agreement is partial and could hold either with what is asserted or inferred from *u*’s utterance, or it can hold with an inferred intention or expectation behind the previous utterance’s goal. Similarly, what is objected to could be partial, and could involve either the utterance itself or inferences/expectations behind the utterance, or it could involve a consequence of the previous turn’s goals (TOD), in which case one needs to infer the speaker’s plan and distinguish problematic consequences or right-hand-sides of planning rules. Here is an analysis of this example roughly following the annotation procedure laid out in the previous section:

1. This looks like concession, with the TC as the plan/goal the CPs are trying to achieve:
   u: [duration(Something) = 2hrs] $\rightarrow$ [total_duration(plan_so_far) = 7hrs]
   s: acknowledge or accept or accept-part
   s: inform(u, precondition[duration(load(oranges)) = 1hr]) $>$ $\neg$ [total_duration(plan_so_far) = 7hrs]

   (a) propositions(X):

      i. duration(B) = 2hrs

ii. so(A, [total_duration(C) = 7hrs])

propositions(Y):

i. yes

ii. but(Z[15], [duration(load(oranges)) = 1hr])

(b) Conventional_implicature(s):[16]

i. $result_{trigger=\text{“so”}}(A_X, [total\_duration(C) = 7hrs])$

ii. $contrast_{trigger=\text{“but”}}(Z_X, [duration(load(oranges)) = 1hr])$

Conversational, i.e., scalar_quantity_implicature(s):

i. $total\_duration(overall\_task) \geq 7hrs$

(c) i. all assertions

ii. Old_Information:

- X: [duration(B) = 2hrs]
- Y: none

New_Information:

- X: [total_duration(C) = 7hrs]
- Y: [duration(load(oranges)) = 1hr]

iii. SA(X):

A. $[verify([duration(B) = 2hrs])$ and/or $assert([duration(B) = 2hrs])$ and/or $thinking\_out\_loud([duration(B) = 2hrs])]$

B. $assert([total\_duration(C) = 7hrs]))$

SA(Y):

A. $accept(W_X)$ or $accept - part(W_X)$ or $acknowledge(W_X)$[17]

B. $assert([duration(load(oranges)) = 1hr])$

iv. Intentions(X):

A. $I1_X = intend[assert([duration(B) = 2hrs])]$

[15]Resolving Z is the goal of this exercise, since doing this determines what is in contrast.

[16]Notice that we assume contrast with something in the immediately preceding turn when we encounter conventional implicatures triggered by “but”, and likewise for “so”. The reason why we use A_X and Z_X here is to indicate that the relation involves something in X.

[17]Where W is a variable representing material from the previous turn that needs to be resolved to determine exactly what parts of the previous turn s agrees with or whether s is simply acknowledging understanding of the previous turn.

B. $I2_X = intend[assert(result_{trigger=\text{“so”}}(A, [total\_duration(C) = 7hrs]))]$

Intentions(Y):

A. $I3_Y = intend[acknowledge(W_X)]$[18]

B. $I4_Y = intend[assert(contrast_{trigger=\text{“but”}}(Z, [duration(load(oranges)) = 1hr]))]$

Intentional_structure:

A. $I5 = dominates(I2_X, I1_X)$

B. $satisfaction\_precedes(I5, I3_Y)$

C. $I6? = dominates(I4_X, I5)$

v. Planning_History (i.e., main goals grounded):

A. u utt9,11: $goal[take(and(1(engine), 2(boxcar)), from(Elmira), to(Dansville))]$

B. s utt13: $effect[previous\_goal, [duration(previous\_goal) = 3hrs]]$

C. u utt14: $goal[go(to(Corning), [reason = pick\_up(oranges)])]$

D. s utt15: $goal[pick\_up(boxcar)]$ and $goal[go(from(Dansville), to(Corning))]$

E. s utt19: $effect[plan\_so\_far, [total\_duration(plan\_so\_far) = 5hrs]]$

F. u utt20,22: $goal[go(to(Bath), from(Corning))]$

G. s utt23: $effect[previous\_goal, [duration(Bath, Corning) = 2hrs]]$

H. u utt24: $effect[plan\_so\_far, [total\_duration(plan\_so\_far) = 7hrs]]$

vi. Planning_Information:

A. X_u: $effect[D, [duration(B) = 2hrs]]$
$effect[plan\_so\_far, [total\_duration(plan\_so\_far) = 7hrs]]$

B. Y_s: $precondition[plan\_so\_far, load(oranges)]$,
$effect[previous\_precondition, [duration(load\_oranges) = 1hr]]$

vii. Topic Under Discussion: duration to finish task

Relations Between (and of) SAs:

A. $R1_X$[19] $= Nonvolitional\_Result_{trigger=\text{“so”}}(SA(X_1), SA(X_2))$

B. $R2_{YX} = acknowledge$[20]$(R1_X)$

C. $R3_{YX} = Contrast_{trigger=\text{“but”}}(Z, SA(Y_2))$

viii. Salient parts:

[18] We address the ambiguity in SA assignment here by using the minimal intention to “acknowledge” rather than the more definite commitments of “agree” or “agree-part”.

A. "so" is a consequence trigger used to infer the Nonvolitional Cause relation.

B. "but" is a contrast trigger used to infer that there is a contrast relation between the clause following the "but" and something that went before; since "yes" immediately precedes "but", we can assume that the contrast lies in something said in the previous turn.

C. "Yes" is used anaphorically here, referring to something across the speaker turn.

2. Summary:

   (a) Resolving unbound variables by unification with instantiated forms:
   $C = plan\_so\_far$ via Planning Information.
   $A = SA(X_1)$ via Relations Between SAs.

   (b) Determining Z: assume that $Z \in X$ following the assumption that X is in contrast with propositions(Y) via the predictions of conventional implicature.

   (c) Evaluate X for the source of contrast:

      i. Prefer new information in X, so assume (until evidence to the contrary) that Z gets bound to $[total\_duration(plan\_so\_far) = 7hrs]$

      ii. Verify this prediction with Planning_History and Planning_Information:
      $effect(precondition(load(oranges)), [duration(load(oranges)) = 1hr])$,
      and $precondition(load(oranges)) \ni Planning\_History$,
      so prediction of conversational (scalar quantity) implicature that
      $total\_duration(plan\_so\_far) \geq 7hrs$ holds,
      since $goal(load(oranges)) \ni plan\_so\_far$.

   So we have resolved Z and determined what is in contrast!

3.2.1.2 Intentions

Intentions in our perspective do not strictly follow Grosz and Sidner's more task-related formulation. While they clearly state that the intentional structure they present is neither identical nor isomorphic to the general plan involved in resolving the task under discussion, they do formulate their intentions in terms of speech and task actions,

goals and expectations. For example, in their flywheel-removal task-solving example dialogue they use nested intentional structures like (Intend Expert (Intend Apprentice (Remove Apprentice flywheel))), (Intend Apprentice (Intend Expert (Identify Expert Apprentice another tool))), (Intend Apprentice (Intend Expert (Tell Expert Apprentice (How (Getoff Apprentice wheel))))) and (Intend Expert (Know-how-to Apprentice (Use Apprentice wheelpuller))). In NTOD, the intentions more often take forms like (Intend Speaker (Intend Hearer (Believe $<$ *proposition* $>$))).

For now we adopt their practice of making propositional information the arguments of intentions, but our emphasis is on the actual speech acts performed and what these acts communicate rather than on the higher level intentions concerning mutual belief and expectations, since we hope that these will arise out of the Information State model of dialogue we will apply. Notice that our SAs take propositional arguments in their first-order logical form, and occasionally also subject-matter rhetorical relations. We also incorporate Kreutel and Matheson's arguments that assertions are bids or offers until they have been grounded as assertions by nesting SAs within *intend* operators, representing the fact that speakers intend to commit various SAs rather than committing them to having committed them before they have been grounded by the other speaker.

3.2.1.3 Types of Contrast in this Example

The assertion of new material in a contrastive construction (we view the clause following "but" to be contrastive because it is an argument of the contrastive cue) is a defeasible signal that the speaker agrees to some of the previous turn's content but disagrees to other aspects of it, or is introducing a new goal. So to restate, the speaker can do one of the following actions given such a "yes but" contrastive TOD construction:

1. contradict directly some of the assertions or inferences launched in the previous turn

2. introduce new goals which have consequences which would not hold given something in the previous turn

3. introduce a new goal that follows in sequence, in which case the contrast is often at a meta-level, defeating the previous speaker's assumption that the task is finished or some goal has been achieved

Recall that the "but" does not deny the whole previous turn (following Grice (1975)) and acknowledges the preceding turn, often agreeing (explicitly or implicitly) to at least part of the previous turn.

3.2.1.4 Summary of our Analysis

The break-down of the previous example distinguishes between many different aspects that contribute to its interpretation. The question to be asked at the end of this exercise is what the overall interpretation/relation to be assigned should be. It is an open question as to whether a conjunction of features (like Knott's feature-theoretic perspective of cue-phrase semantics (Knott, 1996)) could best serve to describe this mini-dialogue without confusing what is essentially going on, or whether it is better to try and isolate a few key features that best characterise what is going on.

In the previous example identification of planning operators communicated and the respective task-plans involved as well as implicature (both conventional and conversational) drawn from the planning information distinguished the most salient features of this example, which required inferring that the total duration of the whole task would take longer than 7 hours, defeating a precondition presented by the user.

Clearly somewhere in the above analysis a representation of the CPs' (Conversational Participants') plans will become necessary to verify our suppositions.

3.2.1.5 A Brief Analysis of TRAINS and TOD

So far the clues we have used to detect cross-speaker contrast when the "but" occurs medially rather than turn-initially are:

1. Anaphoric/deictic reference in the first argument of the contrast to something said in the preceding turn, e.g. "that" in utt42 refers to *s*' assertion (paraphrased from d93-20.2; see Appendix A for the analysis):

```
utt40 : s:  [Elmira to Bath is four hours]
utt41 : u:  [it's four hours]
            [so <two three> so we can get that there by seven in the
            morning]
utt42 :     [[then that's not a problem]] BUT [[I don't see how I can
            get the boxcar of bananas and the boxcar of oranges
            to Bath by twelve if we gotta go all over the place to
            pick up the boxcars]]
```

2. Restatement of something the other speaker said.

3. A representation of the evolving (meta-level) plan being formed, along with information about who contributed which parts. We assume that contrast with the evolving plan is an example of cross-speaker contrast if the first argument of the contrast (i.e., the part of the plan being objected to) was contributed by the other speaker.

4. Also if "but" precedes a question we can assume it is cueing cross-speaker contrast.

The types of cross-speaker "but" signalled contrast we have seen so far in TOD are:

1. concession (will be explored more in Chapter 6)

2. objection to an aspect of the evolving plan that is been contributed by the other speaker

3. correction (in case of misunderstanding or misassumption that things were grounded earlier, explored more in Chapter 7)

3.2.2 The Switchboard Corpus

As mentioned earlier, the Switchboard corpus involves two strangers given a nontask-oriented topic to discuss over the telephone. As we mentioned earlier, in the Switchboard corpus "but" appears utterance-initially about 23.6% of the time when it appears

(excluding cases where it follows phrases, false starts or single-word agreements or disagreements like “yeah, but”), which indicates that a decent proportion of “but”s relate material across turns in this NTOD corpus. In many of the dialogues, one speaker takes the initiative and recounts a story or makes an argument, while the other speaker mostly backchannels, so we see a lot of same-speaker contrast across turns, as in the example below.[21]

```
SPEAKER_B: so i was very comfortable you know in doing it when it got
           to the point that we had to do it but there's well i had an
           occasion for my uh mother-in-law who
SPEAKER_B: had fell and needed to be you know could not take care of
           herself anymore was confined to a nursing home for a while
           that was really not a very good experience uh
SPEAKER_B: it had to be done in a hurry i mean we didn't have you know
           like six months to check all of these places out
SPEAKER_B: and it was really not not very good uh
SPEAKER_B: deal we were not really happy with the
SPEAKER_A: yeah
SPEAKER_B: nursing home that we finally had fortunately she only had
           to stay a few weeks and she was able to to return to her
           apartment again
SPEAKER_B: BUT it's really a big uh big decision as to you know when
           to do it
```

Many of the Switchboard examples involve speakers holding essentially long monologues, with the other speaker back-channelling occasionally as above. This is not the sort of dialogue we are really interested in analysing, as it is essentially monologue.

Our analysis of TOD was heavily based on the idea of an evolving task-plan that speakers are forming which coincides with their often mutually-exclusive goals, plans and beliefs. Perhaps the analogy in NTOD to a shared plan is a shared Common

[21] All Switchboard examples can be found off the LDC web-site at
http://www.ldc.upenn.edu/cgi-bin/lol/swb/viewdoc?corpus=swb &datatype=java&javawidth=600
&javaheight=200&word=but&wtype=raw&index=region&position=
followed by a given index which will be specifi ed for each example. (For the example above, this index is 2537#MATCH, which is appended to the url above.)

Ground which contains the evolving topic, with a representation of what each CP's opinions and beliefs are.

Here is an example of cross-turn "but" in Switchboard; notice that the difference between NTOD and TOD here is that what is being conceded is explicit in the dialogue itself, and there is no need to represent meta-level plans, beliefs, goals or conflicts to resolve relations in NTOD in this case: (this paraphrased example also comes from the same dialogue as above)

```
SPEAKER_B: [yeah]
           [[just because they're grandparents doesn't automatically
            make them a good child carer]]
SPEAKER_A: [yeah] BUT [[i've had a lot of good experiences with many
           people especially where they've had extended family]
           and [i see that we may need to like get close to the family
           environment and get down to the values]]
```

Analysing this example via an abbreviated version of the annotation procedure presented earlier we have:

1. (a) X: $because[grandparents, not(make[automatically, grandparents, good(child\_carer)])]$
   Y_1: $had[good\_experience\_with(B), A] IF had[extended\_family(B)]$
   Y_2: $need[we, to\_get\_close\_to(family)]$

   (b) X_{infer}: $\neg[being\_grandparents(X) \rightarrow good\_childcarer(X)]$; i.e., X denies that grandparents are necessarily always good childcarers. Note that X does not deny that they can sometimes be good childcarers, she simply wants to deny the entailment; possibly she would not mind a defeasible implication of the same rule.
   Y_{infer}: $extended\_family(X) > close\_family(Y)$ AND $good\_values(Y)$, i.e., extended family (presumably as childcarers) tends to result in a situation where the family is close and there are good values

   (c) i. X and Y relate assertions

   ii. X_{old}: $are(grandparents, they)$

   iii. SA_X: $assert(X)$
   SA_Y: $assert(Y1)$, $assert(Y2)$

iv. $counter(Y_2, X_{infer})$

v. Topic Under Discussion: whether grandparents are good childcarers or not.
Y_{rhet}: $reason(Y_2, Y_1)$; $counterevidence(Y_2, X_{infer})$; $agree(Y_1, X)$

vi. trigger: "but" in Y; "yeah" in Y

(d) Summary:
Y_1 and Y_2 provide evidence that $being\_grandparents(X) \rightarrow good\_childcarer(X)$, lending strength to the rule as opposed to X_{infer}, which argues against assuming this expectation

The annotation gives useful information about the nature of the contrast, namely that there is some claim being argued for by X framed as an inferred rule which Y refutes by presenting counter-evidence in the form of defeasible rules arguing that grandparents (since they are a subgroup of extended family) can be good childcarers, since extended family caring for children result in a close family with good values. Thus Y argues that they are good childcarers while not denying X's claim that they are not always good childcarers. We will call examples like this cross-speaker concession, and will present a treatment of concession and discuss cases like this more extensively in Chapter 6.

NTOD displays a greater range of relations than TOD. In the example below we see DofE[22] (paraphrased):

```
SPEAKER_A: [[it changes with the weather]]
           [wear light clothing if it's hot out]
SPEAKER_B: [yeah] BUT [[then usually in the summer it's cold in
           the offices because the air conditioner's doing its job
           so well]]
```

Here A introduces the rule $[hot\_weather] > [wear\_light\_clothes]$, and B denies the expectation by saying $[it's\_summer] > [cold\_in\_offices]$ and $[cold\_in\_office] > [wear\_warm\_clothes]$ which provides a causal explanation that $reason[cold\_in\_office, (a.c.\_on\_AND\_a.c.\_working\_well)]$. Essentially B communicates the defeasible expectation that $[it's\_summer \& hot\_weather] > [a.c.\_on]$ and $[a.c.\_on] > [cold\_in\_office]$. This case involves DofE, although it is a bit

[22] index 189#MATCH

difficult to see since the antecedent of the rule launches another rule which has results that are contradictory to the consequent of the first rule.

Below we have an example of meta-level contrast in Switchboard. The appearance and frequency of meta-level contrast indicates that we will have to consider how "but" relations work on a SA level as well as on a strictly semantic level. It is not easy to systematise thinking about contrast at the SA level, since these cases often involve putting forward a different perspective on a topic, either at the level of topic (i.e., a semantically contrastive argument which is related to a meta-level argument) or at the level of topic management as seen below[23] (paraphrased):

```
SPEAKER_A: [[i believe we've pretty much summed everything up]]
SPEAKER_B: [i know]
           BUT [[i remember you talked about something you
           started off]
           and [you talked about the telephone calls]
```

3.2.2.1 A brief analysis of SWITCHBOARD and NTOD

Switchboard examples often need considerable prior context (sometimes even the first few turns of the dialogue) to get a sense of what the contrast involves. Also, contrast is often made with something the same speaker said earlier. This seems to happen a lot in dialogues in which one speaker has much more initiative than the other and controls the topic. On a different note, cross-speaker DofE seems much more frequent here than in TOD.

The downsides of Switchboard and possibly also most NTOD corpora include:

1. a lack of mixed-initiative within a discussion seems fairly common ($\geq$ 30% at a rough guess)–although as noted above, perhaps the analogy in NTOD to a shared plan in TOD is the evolving shared topic, so that some of these same-speaker "but"s might turn out to be cross-speaker cases.

2. sometimes it is difficult to figure out what the speakers are talking about and to extract what is in contrast semantically from all the extraneous junk.

[23] index 18761#MATCH

The positive aspects of Switchboard (and possibly extendable to other NTOD corpora) include:

1. many more (and varied) rhetorical relations
2. arguments to relations are more linguistically explicit than in TOD where they are most often part of plan and therefore deduced rather than explicitly linguistically expressed in the dialogue
3. interesting (but rare) case of cross-speaker contrast where the ”but” speaker simply introduces the contrast relation by applying it to arguments the other speaker has given but not necessarily contrasted; this is simply a restatement, where the speaker emphasises contrast between the restated arguments.
4. another interesting facet to NTOD is that contrast can involve opinion; while this could happen in TOD too, it is probably more common in NTOD.

Relations seen in both TOD and NTOD include:

1. concession–often signalled by ”yeah, but __”.
2. denial of expectation (DofE)
3. corrections

Some Conclusions:

One of the first findings was that syntactic information does not help much at all to distinguish cross-turn “but”-cued relations in dialogue. Probably this is due to the less formal and real-time aspects of dialogue; speakers often digress, backchannel, interrupt themselves and turns can often consist of a single word or phrase rather than a complete sentence. Here we will wrap up some of our findings from this chapter.

The goal here was to motivate this study by presenting and analysing in detail examples from the corpora. Unfortunately there were not many examples from the corpora involving cross-turn “but”, and those that we found often involved bridging inferences, inferring implicatures (as in the TRAINS example presented earlier), resolving what aspects of the speakers’ task-plans were being referred to, and other inferences that are outside the scope of this work. Given the sparcity of data that involves

cross-turn “but” in dialogue corpora and the lack of automated tools to either isolate these cross-turn examples or perform this sort of in-depth analysis of them, we cannot present statistical evidence of trends across the board, and the examples presented in this chapter should not be taken to be indicative of general trends. For this reason, most of the examples considered in subsequent chapters in this thesis are handmade rather than from the corpora. The goal behind the material presented in this chapter is to learn something of how “but” behaves in real corpora, and to present an analysis on multiple levels of these examples in order to gain insights into how “but” communicates contrast across turns in dialogue.

Possibly the most significant information that comes out of looking at the data is that there are no predictable syntactic constructions in spoken dialogue that give clues about what is in contrast in cross-speaker “but” cases, and in both TOD and NTOD the contrast is often meta-level, involving reference to plans in TOD and to SAs and inferences in NTOD. The results of annotating and thereby analysing the examples from TOD and NTOD corpora also provide important insights that will be kept in mind when modelling the DofE, concession and correction relations formalised in the rest of this thesis.

Chapter 4

Distinguishing Relations

In this chapter we build upon treatments of "but" in monologue to address the relations it conveys across turns in dialogue. This work addresses coherence of dialogue (i.e., how subsequent speaker turns are related) when speakers indicate contrast cued by "but". Here we will focus on how we can distinguish between Denial of Expectation (DofE), concession, Semantic Opposition (SO), correction and rejection. We will start by introducing the relations and then present here a core unified logical description of the phenomena which distinguishes the various kinds of relations.

Motivation Identifying how turns in dialogue are related establishes how they are coherent, and enables more responsive generation that can directly contest these relations themselves; e.g., in DofE, if A can infer B's underlying expectation she can respond directly. For SO while there are no defeasible expectations involved, if we know that B is giving an alternative perspective, then we can respond to this contrasting assertion directly, adding a level of interpretation above the propositional content. Determining what is communicated implicitly in these "but" cases involves determining what relation the "but" is cueing and any accompanying inferences or expectations. We will start by presenting the relations before turning to how they can be logically formulated.

4.1 Relations Signalled By "But"

Medial "but" in monologue licenses DofE, concession and SO, as first put forth by Lakoff (1971). As we will see in dialogue across turns, turn-initial[1] "but" additionally signals rejections and corrections which are dialogue-only phenomena. Furthermore, we will see here that in dialogue, concession, DofE and SO communicate quite different information than their monologue counterparts.

4.1.1 Denial of Expectation

DofE and concession both involve entailment (triggered by "but") of (defeasible) expectations; in the case of DofE, the expectation is then found not to hold in the second clause (Lagerwerf (1998)). We argue that these defeasible expectations can be extended to hold across turns in dialogue, as in the example below:

(4.1) A: Greta Garbo was beautiful.
B: But she never married

Here B has the defeasible expectation that beautiful people usually marry, i.e., $beautiful(X) > married(X)$, via defeasible implication ($>$), which is triggered by interpreting A's assertion that Greta was beautiful ($beautiful(greta)$), combined with the knowledge that she did not marry ($\neg[married(greta)]$), thereby denying the expectation. B can be implicitly either agreeing or disagreeing with A's assertion that Greta was beautiful; if she agrees, then she indicates the surprising DofE that Greta did not marry. If she disagrees, then she expresses the DofE in order to introduce evidence that is contrary to A's claim. Although we do not address the effects of Information Structure (as discussed by Kruijff-Korbayová and Webber (2001) in monologue) in this work, prosodic information could be used to distinguish between these two possibilities; e.g., strong stress on "married" correlates more with the disagreement reading. We will not address prosodic effects in this approach.

[1] I.e., preceding the main clause of the turn, since it can follow "Yes", "OK", etc. and still relate material in the current turn with earlier utterances, as will be seen in upcoming examples.

4.1.2 Concession

We follow the treatment of concession put forth in monologue by Lagerwerf (1998) entailing defeasible rules from the related clauses favouring and disfavouring a contextually available claim or *tercium comparationis* (TC). In dialogue the turns separated by "but" can be evaluated with respect to the TC, and in Task-Oriented Dialogue (TOD) we can evaluate speaker's perspective with respect to the TC via their task-plans in the IS. E.g., examples like the one below can be seen as concessive:

(4.2)
B1: What should we add next?
A2: Let's add the mushrooms.
B3: But I thought we need to add the beans first.

Here A2 favors adding the mushrooms and B3 proposes an alternative. B acknowledges[2] A's proposal but indicates an alternative course of action with respect to the question under discussion (B1) which serves as the TC. So the concession expresses acknowledgement of the assertion but indicates that the assertion fails to answer the question under discussion. Determining where the contrast lies requires accessing the speakers' planning information in order to detect discrepancies and facilitate alignment of their task goals.

Is this really argumentative? However, thinking of the two turns as supporting and denying the TC does not add much useful information here and may not even be accurate, since the speakers are not disputing a claim but resolving a confusion; it might be more useful instead to think of the two turns as expressing different perspectives on the environment. So while argumentative stance *per se* is not an issue here, determining where the discrepancy lies is crucial, and viewing A2 and B3 as conveying different goals with respect to the next step in their joint plan resolves confusion. In other words, determining the TC and the speakers' perspective with respect to this TC facilitates this realignment by isolating the discrepancy in their plans and establishing how the proposals raised in A2 and B3 both need to be accounted for in their joint plan.

[2]We will minimally assume that B acknowledges A's proposal to avoid committing ourselves to either assuming B's acceptance or disagreement/rejection of A's proposal in this example.

4.1.3 Semantic Opposition

Semantic opposition (SO) is a contrastive semantic relation that does not entail defeasible expectations which are then denied like DofE and concession. SO arises via the parallel syntactic structure of the contrasted clauses, and Spooren (1989) argued that the contrast is introduced by incompatible predicates, e.g., "Bob is tall but Mike is short", or "Bob likes football but Mike hates it". According to Spooren, SO is about two entities in contrast in the domain of conversation. One can imagine cases like this in dialogue, where A utters the first clause and B the second. Consider B3 in Example 4.3 below; while this has the parallel syntactic structure of Spooren's definition, the "but" clause is uttered by a different speaker. This enables concessive interpretation, where one might argue that Ingrid's being married is used to make a counter-argument to some claim (e.g., that moviestars are often single) which the former speaker was trying to argue for. The difference lies in whether such a salient claim (or TC) is being debated about or not. Certainly in both cases simple contrast also holds.

4.1.4 Rejection

Across turns in dialogue, seemingly self-contradicting examples in monologue like "Greta was single but married" work marginally better because they are communicated by different speakers (see B1a in the example below), although they are definitely still odd when connected by "but". Notice how much more natural the directly negated B1b below seems as rejection. One reason that "but" does not work well as a marker of rejection is because logically it is identical to "and", and the sense of contrast arises via the *conventional implicature* it is associated with (Grice, 1975). We will define cases like B1b as rejection. Since it is odd when marked with "but", we will not address rejection in this thesis.

(4.3)

A:	Greta was single.
B1a:	? But she married. *rejection?*
B1b:	No. She was married. *rejection*
B2:	But she married in '49. *correction*
B3:	But Ingrid married. *SO/concession*

4.1.5 Correction

In the example above, B2 differs from B1a-b in that it introduces new information to add evidence for the assertion that she married beyond simply rejecting A's assertion as in B1a-b. Continuing to explore the parallel structure of SO, we have seen how rejection differs from SO, and here we will focus on a type of correction which also involves contrasting predicates. In B2 above, it is possible that B accepts A's assertion at a time before 1949, and introduces new information that invalidates A's claim in '49 and after. It is also possible that B2 implicitly denies A's assertion. It appears that in these parallel syntactic structure cases, we can distinguish rejections from corrections as involving explicit denial rather than implicit denial, as in the case of correction. As we saw in Bb, explicit denial involves negating A's turn, which is still possible for correction provided that new information which is mutually exclusive to A's assertion is also asserted. However explicit rejection, as we saw in Ba, is conveyed best without "but" and with direct negation instead. A last point to note is that in SO, correction and rejection, what is contrasted is at the semantic level and does not involve entailment as in DofE and concession.

4.2 Logical Formulation Of Cross-Turn "But"

In this work, we will follow Grote *et al.* (1995) in arguing that a unified scheme can be used for DofE, concession, correction and SO, where the difference arises from the variable which is realised. I.e., Grote et al. argue that the following general situation relating propositions and implications holds (in monologue):

- $A \rightarrow C$
- $B \rightarrow \neg C$

We adapt Grote et al.'s basic ABC-scheme for concession across turns, where C stands for the TC. (We will replace their A, B, C with p, q, r respectively in the discussion below.) This underlying formalism can be extended to address SO, rejection

and correction as well, giving us a very handy framework for analysing many different sorts of "but"-signalled contrast across turns. We will argue for four alternative situations based on Case 1 below.

4.2.1 Case 1: Concession

Given the basic scheme proposed by Grote et al. adapted for cross-turn contrast, and verbalizing p and q, we get a situation that is typically seen as concessive (where B has the expectations that $p > r$ and $q > \neg r$):

(4.4) A: p
B: But q
$p > r, q > \neg r$

Considering the example below, Kruijff-Korbayová and Webber (2001) argue that its monologue version (i.e., "Although he does not have a car, he has a bike") involves concession[3]. In dialogue, the TC could be that the subject either has or does not have a means of transport:

(4.5) A: He doesn't have a car. p
B: But he has a bike. q
$r =$ he doesn't have transport/he's not mobile

While this example bears a superficial resemblance to SO due to its parallel structure (Asher (1993)), we will analyse it as concession here, since there are only two elements which differ, i.e., mode of transport and whether the subject (which is in common) either has or does not have the given vehicle, which lends itself nicely to the argumentative concessive interpretation. The concessive interpretation poses two defeasible rules (i.e., $p > r$ and $q > \neg r$ from B's perspective). Recall that Spooren (1989) argued that SO arises from incompatible predicates, which in monologue requires different subjects in order to avoid contradictions like "Greta is single but married". However as we saw earlier, the contradiction is removed when the opposing clauses span turns in dialogue. Indeed the dialogue version functioned as correction or rejection (as in Example 4.3).

[3] Korbayova and Webber call "concessive opposition" what we call concession; they define concession as further distinguished into concessive opposition and DofE.

We will distinguish cases like Example 4.5 above from SO/correction/rejection on the basis that the former cases do not involve contradictory predicates (e.g., being single vs. married as in Example 4.3), and instead involve alternatives, as discussed for monologue in Kruijff-Korbayová and Webber (2001). I.e., if the dialogue involves the same subject in p and q, and the predicates attributed to this subject are alternatives rather than mutually exclusive states, and additionally we can interpret $p > r$ and $q > \neg r$ from B's perspective, then we will argue that the case is concessive.

In the case of SO, no expectations are entailed by p and q with respect to any contextually available r, and p and q involve contrasting predication of the same subject, so SO is distinct from concession and DofE (as will be seen below), but is not readily characterizable by the p,q,r scheme due to its lack of entailed expectations. Rather we can examine the turns in order to check for the common subject and mutually exclusive predicates that we argue define SO (or rather its analogues correction and rejection) across turns in dialogue.

Lascarides and Asher (2002) argue in the example "John bought an apartment but he rented it", that the clauses are related by both Narration and Contrast relations, where the correct temporal effects arise via the Narration relation. However their Contrast relation does not capture the idea that the first clause entails the expectation that one buys an apartment to live in it, which is denied by the second clause. I.e., it does not address argumentative or expectational aspects of "but" (i.e., concessive interpretation) or (in this case), the denied expectations entailed by "but". stretch to see how the Contrast relation proposed by Asher (1993) seems closer to the definition of SO put forth in Lakoff (1971). For example, consider the examples below, which do not involve isomorphic structures and require bridging inferences to determine how the turns are related. These examples are much harder to consider as SO, with p,q,r shown to illustrate how they map onto the basic distinction scheme advocated:

(4.6)
A: Bob intended to go. p
B: But he had visitors. q
r = Bob went.

(4.7)
A: Bob went to the church. p
B: But the vicar wasn't there. q
r = Bob spoke to the vicar.

Following Asher, these two examples do not meet the parallel structure criteria of Contrast. Although SDRT does not address concession, this seems a much better analysis of Example 4.6, with r representing the TC. In both cases, $p > r$ and $q > \neg r$, so Example 4.7 can also be interpreted as concession with the given TC. However we can also interpret Example 4.7 as DofE with the expectation that *going_to_the_church* $>$ *meeting_the_vicar*, so it has multiple possible interpretations. This is not problematic however, since many of these examples can be interpreted in a few different ways, where context enables preference of one interpretation over another. Distinguishing DofE, concession and SO also depends on what is communicated in each turn, what the speakers believe, and what the topic of discussion involves. However when p, q and r are all distinct, and A communicates p and B q, if B has the expectation that $p > r$ and $q > \neg r$, where r is not identical to either p or q, concession will often be the best interpretation.

4.2.2 Case 2: Denial of Expectation (DofE)

DofE can be seen as a special case of concession, following Lagerwerf (1998) and Lakoff (1971), which typically involves an implicit expectation which can be formulated as a defeasible rule which is then violated (Lagerwerf (1998)). We will describe situations like the one below as involving DofE, since q is not verbalised at all, and we simply have the denied consequent $\neg r$ of the defeasible expectation $p > r$.

(4.8)
A: p
B: But $\neg r$
$p > r$

For example:

(4.9)
A: Bob stayed up all night with the baby. p
B: But he looked fresh as a daisy. $\neg r$
$r =$ Bob looked tired.
$p > r$ staying up all night with the baby $>$ looking tired

Alternatively, we could describe this as:

(4.10)
A: $p > r$
B: $q = \neg r$, so $p > \neg q$

This is typical DofE, e.g.:

(4.11) A: Mary looked all over the shop to find miso sauce. $p > r$
B: But she didn't find any. $\neg r$
$p > r$ looking all over the shop $>$ finding miso sauce

(4.12) A: Bill took the lid off the pot. p
B: But there was nothing inside. $\neg r$
$p > r$ taking the lid off the pot $>$ finding food inside

Example 4.9 above involves cause and effect; normally staying up all night means one looks tired, but (surprisingly) Bob does not look tired. Example 4.11 is a bit odd since the expectation (i.e., the defeasible rule) is explicitly given in A's turn rather than inferred to be in B's private beliefs. This means that B accepts A's expectation and simply denies the consequent, because either the expectation is already in B's private beliefs, or B adopts A's expectation.

Example 4.12 also seems strange when viewed as $p > r$ since p is a precondition or action which needs to be performed in order to achieve the goal, which is to be in a state in which r holds. This is similar to Example 4.11 above which Knott (1999b) calls *plan-based*. In his thesis (Knott (1996)), he distinguishes cases like Examples 4.11 and 4.12 as *result-driven* as opposed to Example 4.9, which is *cause-driven*. We view examples like these as closely related to TOD examples involving agents performing a task together. In TOD, we often see situations which can be mapped onto planning operators in the speakers' task-plans where the expectations arise via the task-plan (e.g., effects of actions, like adding vinegar causing tanginess below), or via satisfaction-precedence (s.p.) between successive actions in the plan. Notice that both speakers need to either (1) be able to reach the same expectations/share the same goals or (2) already share the same expectations/goals.

(4.13) A: I added the mushrooms. p
B: But we need to add the beans first. q

Notice here that B has a different temporal ordering of events in mind than A. These events are ordered by s.p. in B's plan; she has adding beans s.p. adding mushrooms, while A either (1) has the same s.p. relation, but has forgotten to add beans, (2) does not have adding beans in her plan, or (3) has a different ordering of ingredients to be

added (so that she thinks mushrooms need to be added before beans). S.p. involves temporal ordering and the notion of agency, such that certain actions/states need to be accomplished/achieved before other ones. However in NTOD we often saw cases with no apparent temporal ordering, like Example 4.3, which simply seems to link two contingent states, that of being beautiful and being married and indicates that beautiful people tend to also be married. Despite the apparent lack of ordering involved, we will not assume that the converse holds and the implication is bidirectional unless the speaker communicates this, since despite the lack of apparent ordering and statement of apparently simple contingency, B does not necessarily believe that the converse also holds (e.g., she might not expect married people to be beautiful). For bidirectionally contingent cases, $r > p$ also. That is, the expectation is bidirectional between p and r. This is usually because they are simply contingent states, and one can not be said to cause the other, be the result of the other, or temporally follow or precede the other. If the situation is causal (or resultative) or involves temporal ordering, then the expectation will not involve bidirectional implication, and the converse cannot be assumed to hold, so $\neg[r > p]$. The example below can be said to be bidirectional, assuming that if one is good at maths then one is usually good at engineering, and vice versa, i.e., the converse also holds given the original ordering:

(4.14) A: Sue is good at maths.
B: But she's not good at engineering.

As with concession, coherently uttering DofE requires assuming that the expectation involved is in the speakers' common ground. For DofE it is unlikely that the expectation was explicitly grounded in the preceding dialogue, and it is more likely an assumption of the speaker of the DofE that the other speaker also shares this expectation in order to coherently utter DofE. For concession, the expectation needs to have been raised or made salient. So we follow Kruijff-Korbayová and Webber (2001) and others who argue that the TC or expectation is presupposed for concession and DofE; the TC/expectation is presupposed by the speaker of the "but", but as seen in responses to DofE like "But beautiful people don't have to marry" (in the A:"Greta was beautiful", B:"But she never married" example), this presupposition does not have to be shared by both speakers. The speaker of the DofE presupposes that the expectation is

in the common ground, but as this response shows, it does not have to be grounded.

In our approach, we will model the expectation as being in the speaker of the DofE's beliefs as a shared belief. Prior to its utterance it may not have been grounded, but it is assumed to be shared. We leave for future investigation the possibility that the speaker of the DofE is not being cooperative, and utters the "but"-utterance regardless of whether the presupposed expectation is shared. For concession, the TC is a topic under discussion and more likely to be in the common ground, but it is unlikely for it to have been explicitly grounded, as the TC generally arises from the immediate dialogue as a salient claim which is being debated. It is not denied (as this would be DofE) but an alternative perspective toward the TC is provided in the "but" turn. This will be covered in greater depth in Chapter 6.

4.2.3 Case 3: Denial/Rejection

Denial and rejection in dialogue are frequent, but are not signalled by "but". "But" should not be able to license rejection, following Grice (1975) and Lakoff (1971), since it is logically conjunction and only conventionally implicates contrast. Indeed in monologue, medial "but" cannot be used to deny the first clause, as in the example "Jim is a bachelor but he's not, he's married to Sue". This example does not make sense in monologue, but it is fine in dialogue across turns, since speakers are entitled to have different opinions and beliefs.

(4.15)
A: Jim is a bachelor.
B: But he's not! He's married to Sue.

The scheme here is:

(4.16)
A: $p = r$
B: $q = \neg r$

(4.17)
A: Jim seems to be balding.
B: But he's not!

We will use the terms denial and rejection interchangeably, although denial communicates a rejection of the truth of what is denied, while rejection often involves turning down an offer or proposal, and occurs much more often in TOD, while denial seems more common in NTOD. The denials and rejections given here are distinct

from the rejections and corrections involving parallel structure that are closely related to SO as discussed earlier. These do not involve parallel structure, and simply involve straightforward denial, expressing negation of A's turn. They are included here because they can also be described via the same p, q, r scheme.

4.2.4 Case 4: Correction

We will focus in greater detail on correction in Chapter 7, where we will consider corrections of assertions, answers to questions, questions themselves and commands. Here we will simply show how corrections like Example 4.3B2 ("But she married in '49") which involve denial of the previous turn's assertion and assertion of salient new information on the topic can also be modelled in the p, q, r scheme:

- A: $p = r$
- B: $q = \neg r$ AND m, where m is new information in the domain of $topic(r)$

Of course in all these cases where A's turn is denied, we assume that lexical antonyms like "single" and "married" are interpreted as such. Also, it is possible to deny/reject a turn in ways besides directly negating/rejecting it or asserting its lexical antonym; for example, an incompatible predicate can be asserted which implicitly denies A's turn, since the situations described in both turns are deemed by B to be incompatible situations in the domain of conversation. In Chapter 7 we will see corrections in TOD which involve effects asserted which preclude the situation described in A's turn, etc.

4.2.5 Case 5: Semantic Opposition

Although we will not address SO further in this thesis, we will specify here how we would treat SO cases like Example 4.3B3 ("But *Ingrid* married") which involve contrasting subjects and predicates where there is some basis for comparison. That is, presumably there is a reason to contrast Ingrid with Greta in terms of their matrimonial state. What is communicated is simply a contrasting predicate attributed to a different subject or object (e.g., "Bob loves pasta but hates bread" is also SO in monologue). Basically two things need to differ: the predicate needs to have its antonym asserted

or be denied itself, and either subject or object needs to contrast. This implies the following scheme for SO:

- A: $a(b,c)$, where a is the predicate, b the subject and c the object
- B: $a'(b',c')$ where one of the following situations hold: $a' \rightarrow \neg a, c' = c, contrast(b,b')$; $a' \rightarrow \neg a, b' = b, contrast(c,c')$; $a = a', contrast(b,b'), contrast(c,c')$

One point to note is that SO does not seem to follow the p,q,r scheme proposed by Grote *et al.* (1995) (as was also noted by Karagjosova (2001)), presumably because there are no expectations inferred. The crucial distinction between SO and correction on the one hand and DofE and concession on the other is that there are no expectations inferred in the former case. We agree with Karagjosova that the distinction hinges on the fact that in SO, no causal expectation can be identified between clauses which is then shown to be defeated, following the claim made by Lagerwerf (1998) that SO is additive and not causal. We will show in Chapter 5 that (for DofE) this expectation need not simply be causal, as previously claimed, but that it might involve a range of different relations which give rise to defeasible rules, excluding only the antonymous relation, as that is how SO arises. Note also that the antonymous relation also does not give rise to expectations beyond one of opposition, and this opposition relation is not denied in SO, instead the opposition relation is the basis for SO.

For SO, *contrast* describes how subjects or objects can differ. Minimally we can say that they have to belong to a set where at least one of their properties is in common. In the "Bob loves pasta" monologue example above, pasta and bread are both foods, and in Example 4.3B3 Greta and Ingrid are both people. Notice that predicates can also be opposed by asserting alternatives and still be termed SO; e.g., "Bob loves pasta but eats bread" and "Bob drinks wine but eats burgers" can also be analysed as SO, but as opposed to cases in which predicates differ by being denied or having their lexical antonym asserted, these more loosely opposing cases can also be interpreted as other relations. For example, the former example could be interpreted concessively, with "Bob loves pasta" giving rise to the expectation that Bob eats pasta, which is then denied by the second clause. The latter example can also be interpreted concessively, with the first clause implicating the expectation that Bob is posh, which is denied by the second clause.

Summary: This chapter presented an underlying logical distinction scheme whereby the different cross-speaker "but" relations can be distinguished. The idea involves extending relations signalled by "but" from monologue to dialogue when the related material spans speaker turns. We saw that the scheme presented by Grote et al. provides a logical framework which we can adapt to model these relations. Notably, the scheme distinguishes DofE from concession by virtue of what is being related; in DofE the two turns are directly related, since the "but" denies the consequent implied by the former turn's utterance. In concession, both turns relate to a third claim in different ways rather than directly to each other. Correction on the other hand involves no inferences w.r.t. either an expectation or TC and denies the prior turn's utterance while providing an alternative or explanation. Understanding what is communicated in these cases and furthermore how these relations relate what has been communicated will be the backbone for modelling these three relations in the upcoming chapters.

Chapter 5

Denial of Expectation

Denial of Expectation (DofE) is a discourse relation which can be signalled by "but" across turns, as we saw in the last chapter. In this chapter we model DofE for both Task-Oriented (TOD) and Nontask-Oriented Dialogue (NTOD) in the PTT (Poesio and Traum, 1998) Information State (IS; Matheson *et al.* (2000)) model of dialogue. Here we focus on DofE in more detail, focusing on the relation underlying the entailed expectation (discussed in the next section) and address how these underlying relations might provide a more specific link between the related information than simply (defeasible) implication ($>$). We will see that these underlying relations across turns in dialogue need not only be causal (as argued previously, e.g., Lagerwerf (1998)) and we will model this information via the update algorithms presented in the IS representing the dialogue state, where this information can be used to facilitate generation of appropriate responses depending on the responder's private beliefs.

Recall the example below from the previous chapter (printed again for convenience). In the case of DofE, B matches A's assertion to the antecedent of a defeasible expectation (in the form of an inferred rule), but knows information (asserted in B's turn) that contradicts the expected outcome of the rule (i.e., she knows that the rule fails in this case). This analysis extends the treatment of concessive and contrastive markers in monologue presented in Lagerwerf (1998) to cross-speaker relations in dialogue.

(5.1) A: Greta Garbo was beautiful.
B: But she never married.

Here we assumed that B has the defeasible expectation that beautiful people usu-

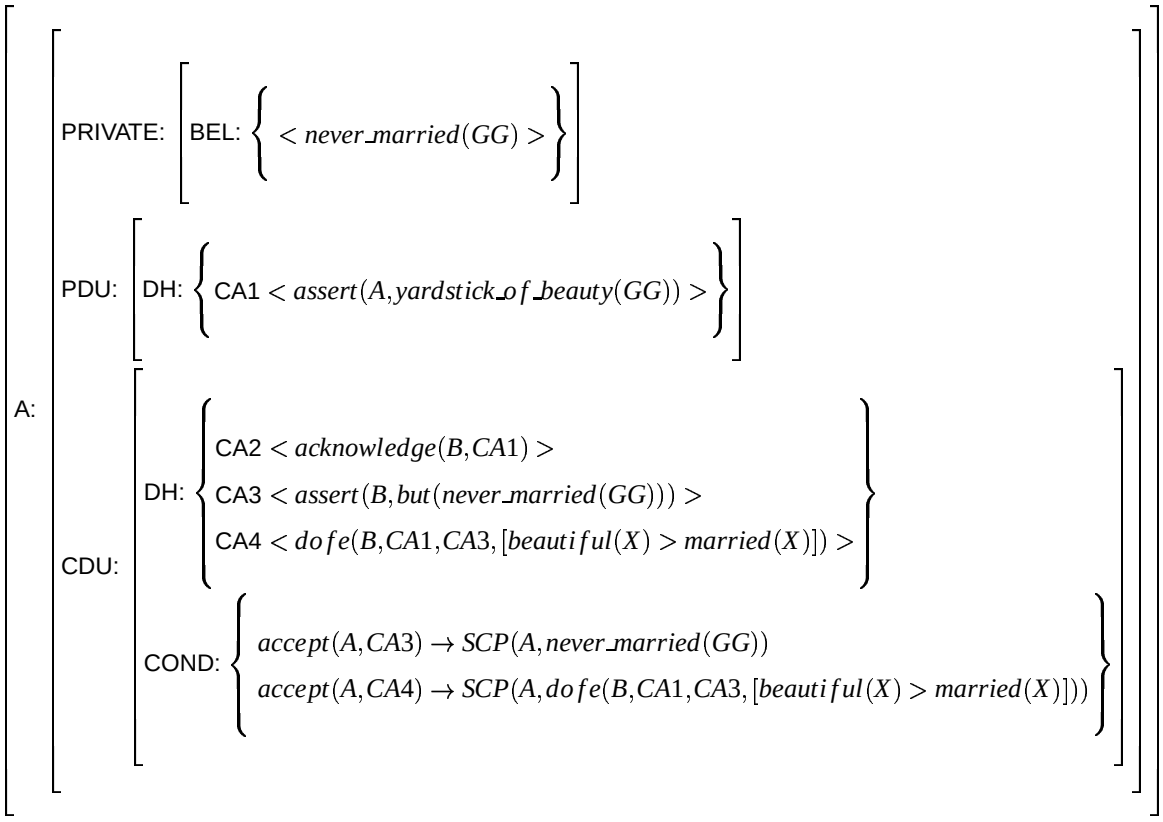

Figure 5.1: A's IS after interpreting B's DofE for Example 5.1

ally marry (i.e., $beautiful(X) > married(X)$, via defeasible implication), which is triggered by interpreting A's assertion that Greta was beautiful ($beautiful(greta)$), combined with the knowledge that she did not marry ($\neg[married(greta)]$), thereby denying the expectation. We argued (in the previous chapter) that B can either implicitly agree or disagree with A's assertion that Greta was beautiful; if she agrees, then she indicates the surprising DofE that Greta did not marry. If she disagrees, then she expresses the DofE in order to introduce evidence that is contrary to A's claim. This Information State (IS) in 5 represents A's IS after interpreting B's DofE for Example 5.1.[1]

Notice in the figure that A also believes (PRIVATE BEL field) that Greta never married, but does not have the expectation that beautiful people need to marry. Interpreting

[1]The fields in the IS are: PRIVATE BELiefs, Previous Dialogue Unit (PDU), Current Dialogue Unit (CDU), Dialogue History (DH) which contains Conversational Acts (CA), and CONDitions, which holds socially committed propositions (SCP) that the speakers are committed to if they accept the left-hand side of the rule. CAs have been annotated with DAMSL (Allen and Core (1996)) tags here. More information on the IS and its fields can be found in Section 2.2.1 and in Matheson *et al.* (2000).

DofE for B enables A to make B's turn coherent given the dialogue so far. Since A agrees with B's assertion (that Greta never married–Conversational Act (CA) 1 above), but has no expectation that beautiful people marry, A would respond "Yes, but beautiful people do not have to marry", signalling disagreement with the expectation implicit in B's assertion. We assume that given no expectation asserting the contingency of being beautiful and being married, A disagrees with this expectation. We have omitted for now the case that A adopts this expectation herself, since we adhere to simply searching A's beliefs; if we were to adopt a more complex model of speakers' beliefs, we could investigate whether A might adopt B's expectation by determining whether this expectation is consistent with all her beliefs (exhaustively, via theorem-proving), but for now we advocate a simple approach that requires a minimum of computational overhead.

5.1 Denial of Expectation in NTOD

Here we aim to model the semantics of "but" in dialogue, focusing on cases in which it signals DofE across speakers. We present an algorithm that predicts the defeated expectation from the perspective of the hearer of the DofE, and we consider differences between TOD and NTOD. We show how it updates beliefs in the the PTT model of dialogue and can be used to facilitate discourse understanding and natural language generation (NLG). Example 5.1 at the beginning of this chapter is a typical example of DofE in NTOD, and we will trace how this example would be processed in the algorithm presented in the next section. In order to understand what B is trying to communicate in Example 5.1, we need to infer that B has an expectation that beautiful people usually marry.

5.1.1 Interpreting Denied Expectations in NTOD

The dynamic treatment of the semantics of "but" in DofE cases was inspired by Knott's proposal to treat such cases algorithmically (Knott (1999a)). However, while Knott addresses "but" in monologue from a third-party perspective, we focus on modelling cross-speaker "but" from the perspectives of the speaker of the "but" and the hearer

respectively, which results in some significant changes to the algorithm he proposed. We propose the following adaptation of his algorithm, which we implement in the PTT model of dialogue. We show how the Greta Garbo example would be processed at each stage of the procedure, interpreting the DofE just after it has been uttered and then updating the IS model of dialogue with the information that B communicated DofE.

1. Initiate the update process with the two propositions being related by "but"; for now we assume that these are just the two adjacent cross-speaker assertions sandwiching the "but".

2. Take the second proposition (i.e., the clause immediately following the "but") and negate it; this is now the right-hand side of the rule. *So "she never married" becomes "she married"* in Example 5.1.

3. Take the first proposition passed to the algorithm and make this the left-hand side of the rule. *LHS = "Greta Garbo was beautiful"*

4. Generalise both propositions by "un-instantiating" them (i.e., replacing any instantiated variables with variable reference), resolving anaphora as necessary.[2] *So* $LHS = beautiful(X)$*;* $RHS = married(X)$

5. We now have the rule, and we pass it back to the update module:
   Rule: $beautiful(X) > married(X)$*, where* $>$ *indicates defeasible implication.*

6. In the update module this rule and relation are added to A's representation of B's beliefs.

Recognising the defeated expectation in these DofE cases enables us to model the speaker of the DofE's private beliefs in the hearer's interpretation process (i.e., beliefs he attributes to the speaker of the DofE). Modelling the DofE of speakers in dialogue allows us to generate the appropriate response to the DofE utterance based on the hearer's own private beliefs with respect to what he believes the other's beliefs are, as will be presented in more detail in the last chapter as an area for further work. The

[2] Tense and mood can also affect the formation of the rule but were not addressed here.

presence of the expectation and contradicting fact also enables the generation of the DofE in B's turn itself, prior to the interpretation of B's utterance, which will also be discussed in Chapter 9.

Since we predict both (1) that B must have the defeasible expectation in her private beliefs, and (2) that this expectation is defeated by a contradicting proposition uttered by A, we can easily generate DofE in the PTT model of dialogue. In 5.1.1 B contains both the expectation and instantiated fact contradicting the RHS of the rule in her private beliefs, and A's assertion instantiating the LHS of this defeasible rule is in CDU and triggers generation of the DofE if B either (1) accepts that Greta was beautiful or (2) disagrees and wants to give evidence to the contrary, namely that one would have expected her to have married if she were beautiful.

We could isolate such expectations in our model of B's (the speaker of the DofE's) private beliefs and check the contents of the previous speaker turn (stored in the CDU) for contradicting assertions which instantiate the LHS of such rules where the RHS is also instantiated to the same constant. The presence of such a co-occurrence could then trigger the generation of an appropriate DofE. Provided that the set of salient beliefs is small (given the context of the dialogue), then searching for rules and contradicting consequents should be a feasible task incorporated into the update rules in the PTT model of dialogue upon hearing a "but". what sort As discussed in Chapter 9, generating DofE, etc. in the first place (i.e., B's turn in the examples above) involves searching the speaker's private beliefs whenever they hear a new utterance; i.e., this search will be part of update after every new utterance.

5.1.2 Modelling DofE

We model grounding of DofE (and cross-speaker rhetorical relations in general) as conditional elements (akin to the treatment of assertions in Kreutel and Matheson (2000)) in the sense that if they are accepted by the other CP, they become *socially committed propositions* (SCPs in the IS below). However, as we will show, the assumption of the expectation itself is defeasible and can be denied by both the speaker and the hearer. The presence of the lexical item "but" triggers a cognitive process modelled by the algorithm, which then predicts what defeasible expectation might be communicated and

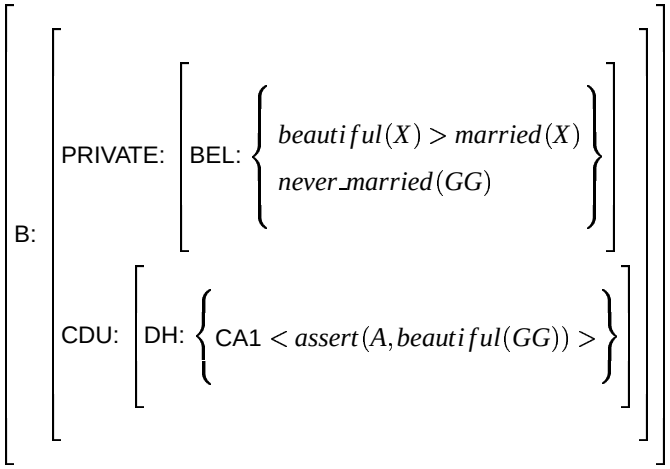

Figure 5.2: B's IS before generating "but" in Example 5.1

defeated.

We model how A might represent B's DofE in the IS for Example 5.1 shown in 5[3] following B's "but" utterance. Since A only has the assertion that Greta never married in his private beliefs, and not necessarily the expectation that beautiful people usually marry, we could expect that A might respond by saying something like "Yes, but beautiful people do not necessarily marry", agreeing to B's assertion but disagreeing with what he assumes to be B's expectation about beautiful people. Although we do not address here other factors that might play a part in determining what A says in response to B, providing no other commitments exist, A has a social obligation to respond to B (Kreutel and Matheson (2000)).We predict how A might respond given her beliefs (after possibly adopting B's expectation), and in Chapter 9 we discuss the different scenarios and how we can generate A's response depending on her beliefs w.r.t. both B's assertion and the inferred expectation attributed to B.

5.1.3 Interpretation Depends on Perspective

We view the procedure or cognitive process described above as predicting that the conventional implicature of contrast attached to the "but" signals the DofE conveyed

[3]Representation of propositions are abbreviated here for brevity, and for the same reason we omit A's representation of B's private beliefs in this IS; hopefully the contents of these will be clear from our explanation.

by B as interpreted by the hearer of the "but" (i.e., A). While the procedure predicts a defeasible expectation that it attributes to the speaker of the "but", we do not check that this expectation is actually in this speaker's private beliefs before predicting the DofE relation. This is intentional, since we model the hearer's interpretation process, rather than modelling the dialogue from the perspective of an omniscient third-party observer. Likewise, in order to generate the DofE, the speaker must have this expectation in her private beliefs. Our assumption of subjective interpretation constrains the hearer of the DofE A to simply predicting DofE for B's "but" utterance, rather than confirming that B is really communicating the relation. So if A is incorrect in his assumption that B is communicating a DofE, then B can correct this misassumption (provided she interprets A's reply as assuming that she is responding DofE). Consider for example the following continuation of Example 5.1:

(5.2)
A: Yeah, but beautiful people don't have to marry.
B: I'm not saying that they do, I'm just surprised that she didn't marry, since so many men were obsessed by her beauty.

Here B reveals a different reason for using "but", namely the expectation that having many men obsessed with one's beauty usually results in one marrying. Here the LHS of the rule is possibly inferred from the "Greta Garbo was beautiful" assertion via a bridging inference. In other words, the procedure does not guarantee that two propositions in adjacent speaker turns linked by "but" are always related via DofE, but merely predicts DofE if the corresponding expectation is found in the speaker's beliefs.

The possibility that A can deny what he assumes to be B's defeasible expectation and that B can deny the same expectation is evidence of the difference in speaker perspective involved. It also shows the utility of inferring underlying expectations. DofE is itself the denial of some defeasible expectation; however, in dialogue the DofE can itself be refuted by both the hearer of the DofE and the speaker herself (if she did not actually communicate DofE).

5.2 Denial of Expectation in TOD

Modelling DofE "but" in TODs can give useful information about the agents' goals, constraints and beliefs to facilitate more collaborative behaviour. Consider the follow-

ing example paraphrased from the TRAINS dialogues (d93-20.2 utt130-1)[4]:

(5.3)
A: so we should be (there) at eleven I think
B: eleven BUT that still doesn't give us enough time to get to Bath

In this example it is possible for cross-speaker "but" to indicate a conflict in planning that does not fit comfortably into the defeated expectation paradigm. While B acknowledges A's assertion of an effect of the plan they are constructing, B communicates a problem with this effect by arguing that it does not allow for a necessary precondition in the next step of the shared plan to hold. Our NTOD procedure models B's expectation as *[be(ing) there at 11] > [gives enough time to get to Bath]* , (i.e., more generally, *[effect of plan so far] > [preconditions for next goal will hold]*), and predicts that B's private beliefs contain both this expectation and the contradictory fact that the effect of the plan (i.e., being there at eleven) does not actually allow for the next goal (i.e., getting to Bath) to be met.

Notice the oddness of such a specific expectation predicted by the procedure, which would also be predicted by Knott's plan-based procedure for "but" (Knott (1999a)). It appears that in this case we need to generalise the expectation in terms of the planning operators the assertions refer to in the joint-plan the speakers are constructing. Rather than relying on stored private beliefs as in the expectation-based Greta Garbo case above, here B actively (1) matches A's assertion to an effect of the plan they are constructing and then (2) checks to see if the next goal in the plan can be achieved. If the goal cannot be achieved, in this case because the effect of the plan so far does not allow the next goal to be met, B signals this with "but", indicating a *frustrated plan* rather than a *defeated expectation*, following Knott. In other words, in this case, "but" launches planning to determine whether the goal can be met, and the planning process returns a negative answer, so "but" signals a frustrated plan.

We distinguish this situation from the defeated expectation sense in two major ways: (1) what is defeated is the accomplishment of a goal in the speakers' joint-plan rather than an expectation on the part of one of the speakers, and (2) determining that this goal cannot be achieved occurs via a planning process launched after hearing A's assertion, rather than searching one's beliefs for a (static) expectation. We incorporate

[4]This example can also be interpreted as cross-turn concession, as will be discussed in Chapter 6.

a call to a planning module in the update process (given A's utterance and B's relevant beliefs) to conduct the planning and return an answer as to whether the speakers' goal can be achieved or not. In the case that the goal cannot be achieved on the basis of A's utterance, a DofE is generated.

5.2.1 Incorporating Planning Information

Planning information can account for the oddness of the highly specific rule predicted by the NTOD procedure (e.g., *[be(ing) there at 11]* $>$ *[gives enough time to get to Bath]*) in TOD examples like Example 5.3. In order to determine how planning information might help resolve some of these problematic TOD cases, we considered simpler dialogues, in which the agents are involved in carrying out the actions they are planning as they speak, rather than forming future plans, as in TRAINS.

We argue that by searching simple task plans, it is possible to detect if a link (usually causal or s.p.) exists between A and B's propositions in the form of precondition/decomposition and effect relationships in the plan. We propose incorporating such a check into the DofE procedure to prevent it from predicting a defeasible expectation directly from the utterances in cases in which the "but" signals a planning problem instead. As seen in the example below, it is possible for B to indicate (1) a problem with a past action in the plan (i.e., a planning mismatch, as in B2), (2) a problem with a resulting effect (e.g., in B3), and (3) a problem with a current or future goal or action (e.g., B1).

(5.4)
A: Add the beans to the sauce
B1: (OK) But we haven't finished sauteeing the onions
B2: (OK) But the beans cook slower than the broccoli and we added that already
B3: (OK) <adds beans> But now the sauce is too thick

In this example, Knott's plan-based "but" procedure would predict the rule [*add beans to sauce* $\wedge$ *unknown preconditions* $>$ *make sauce*] based on A, which should then be defeated by B in a dialogue-version of his plan-based procedure.

However B1 does not indicate whether the rule succeeds or fails, since in fact the outcome of the goal has not been determined yet. B1 could be analysed as introducing a new precondition to adding the beans which A might not have known about. Knott's

procedure also fails to account for the added information about a planning mismatch in B2. B2 indicates that B had a different plan in mind than A, since she had probably assumed that they were not going to add the beans, otherwise presumably she would have indicated that they should add the beans when A suggested adding broccoli.

So B2 is rejecting A's suggestion on the grounds that the overall plan (making sauce) would not succeed if they were to add the beans after the broccoli. B3 does not disagree with A at all, and performs the action, but then finds that the overall goal has not been met successfully, since the sauce is too thick; this case is closest to the procedural analysis of DofE, but conveys extra information about which actions have occurred which is not represented in the simple defeasible planning rule Knott predicts. As for cases like B1 and B2, it is uncertain whether they can be labelled as DofE at all, since they do not seem to involve defeated expectations or rules. None of this information is captured by Knott's plan-based procedure, and predicting simply that the rule fails because the goal is not achieved misses much of the planning information communicated in these cases.

Clearly some part of the planning process, both in terms of matching assertional material in the adjacent speaker turns to planning operators in the plan being constructed and then checking to ensure that the next goal can be met given A's information will be necessary to account for cases like Example 5.3. Accounting for them all as DofEs where the expectations arise directly from the utterances does not necessarily predict sensible rules, and misses much of the additional planning information conveyed.

5.2.2 Semantics or Planning?

While semantic representation systems like Lexical Conceptual Structures (Jackendoff (1972)) or VerbNet (Palmer (1990)) address inferences that arise from actions, neither these approaches nor others that do not make use of contextual information and prior discourse can address the sort of defeasible expectations that arise via rhetorical relations like concession or DofE. Work by Lagerwerf (1998) proposes a lexical semantics for causal connectives like "but" in which they invoke an entailment in the form of a defeasible implication in monologue, which can also account for many NTOD DofE

cases like the Greta Garbo example. However for dialogue the difference in speakers' beliefs needs to be accounted for and represented in the semantics in order to generate relevant and appropriate responses depending on the hearer of the DofE's own beliefs w.r.t. both the expectation and asserted fact that contradicts this expectation. Knott (1999a) addresses the distinction between *expectation* and *plan* based "but" and presents procedures for DofE in monologue which we base our approach on to present an procedure that models DofE in dialogue and addresses dialogue-related issues in the PTT IS framework.

TOD examples like Example 5.5 below involve highly task-related and specific expectations with short "life-spans" during which they are valid, and are not part of speakers' private beliefs as we claim for NTOD, but instead convey expectations about salient (and current) planning operators in the speaker's task plan. E.g., the expectation that "adding mushrooms implies that (usually) beans have already been added", requires access to the task plan and contextual and domain information in order to license interpreting or generating DofE.

Clearly these expectations are not static like beliefs about beautiful people usually marrying, and are only valid given the recipe and current task-stage. We will argue in the chapter on concession that the same holds for concession, where the expectations hold w.r.t. a claim that is being disputed and determining this claim occurs via the task plan and dialogue history, and not through lexical inferences at all, though here we will focus on the presentation of an procedure for DofE in TOD. In TOD examples from the corpora, e.g., the TRAINS example below[5], DofE interpretations are possible but require interpreting the propositions as the planning operations they are indirectly proposing in order to determine what is really being implicated.

(5.5) U: well actually I'll already have an engine in Bath after I unload the boxcars right
S: right but you wouldn't have it in time

So for this example, in order to understand that U is proposing a new goal, namely to go back to Corning with the engine and boxcars in Bath and pick up more oranges, one needs to recognise what is being indirectly proposed, which requires accessing the task plan. Green and Carberry (1992) address this by proposing that discourse

[5]Paraphrased from the TRAINS dialogues (d93-15.2), which can be found at www.cs.rochester/research/trains/.

plan operators can be viewed as defeasible rules expressing typical (normal) effects of illocutionary acts in a context with certain applicability conditions. They model the claim that indirect replies conversationally implicate direct responses by proposing discourse relation inference rules between the question and an indirect reply in question-answering situations.

They propose a discourse plan operator to infer concession which deals with answers such as “no, but I scraped them” in response to the question “did you wash the dishes?” Their operator does not address cross-speaker concession where arguments for and against the claim being debated are in different turns and assumes that the first clause of the relation (e.g., “no” in this case) is a proposition that an agent failed to do an action (of act type) T, and requires that the second clause is a proposition that describes either (1) the satisfaction of an applicability condition of T, (2) a precondition of T, (3) the success of a step of T, or (4) the achievement of a goal of T. If these conditions hold, then the operator predicts that concession between these clauses is plausible.

Our approach has wider scope since it addresses relations across speakers, which entails searching the speaker of the “but”’s plan for task-related intentions and addresses cases in which the first speaker’s proposition often indirectly conveys task-related planning intentions which the second speaker refutes as in the above example. We ascribe defeasible expectations to speakers if they convey DofE following Lagerwerf (1998) and test to see if inferring DofE is applicable via the procedure proposed in section 5.2.7, resulting in a hybrid approach that takes advantage of both semantic and planning approaches within the IS framework to minimise planning overhead and yet model these plan-based expectations.

5.2.3 Interpreting TOD

Indirect SAs in TOD often convey proposed task actions and plans which must be resolved in order to understand what a speaker is trying to communicate. While others have addressed the resolution of indirect SAs (Stone (2000), Cohen and Levesque (1990)), some even in TOD (Maier (1996)), our proposal is novel in two major ways: (1) it aims to resolve DofE across speakers in TOD which is often communicated via

indirect SA and involves interpretation. Also, inferring speakers' expectations facilitates interpretation of subsequent dialogue, and (2) our approach does not require either theorem proving with a logic-based SA theory or the sort of heavy-duty planning involved in Carberry *et al.* (1993) in order to generate the appropriate responses to DofE. They argue that implicit propositions and contextual differences cannot be accounted for by discourse relations alone since interpretation that depends on background knowledge and recognising intentions is essential for dialogue understanding. Their tripartite dialogue model (Lambert and Carberry (1991)) is based on the idea that in order to account for intentions in dialogue, the domain, problem-solving and discourse-level actions all need to be accounted for.

While we agree with this claim, we propose resolving task-plan related indirect SAs via a Conversational Act (CA) interpreter that maps automatically determined direct SAs and their associated propositions in the IS onto planning operators and forward-chains on these operators in the planner to determine what indirect SAs are being communicated in these cases. So our approach does not require the overhead of building full-scale planning models for the task, discourse and domain levels, since we do not make use of domain-level information except where it is part of the relevant stages of the joint-plan used to model the task and planning history. Discourse-level information including SAs, intentions, obligations, and what is already been grounded or awaiting grounding is represented in the IS, along with the speakers' beliefs and dialogue and planning history (DH and TPH in the IS). E.g., consider the following TOD example:

(5.6)
A: Let's add mushrooms to the sauce
A′: I think we've got some mushrooms here.
B: *But* we haven't added beans yet.

This example involves disagreement in temporal ordering of events, or in planning terms, a mismatch in plans, where the action B mentions should precede A's proposal (adding mushrooms) in B's plan. We view the "but" in cases like this as signalling DofE, since inferring the defeasible expectation $adding\_mushrooms > have\_added\_beans$ which is defeated in B's turn gives us just the sort of expectation about ordering that we want to represent in our model of the speaker's plan-related expectations to explain the coherence of "but" in TOD cases like this where opposing goals (e.g., adding

mushrooms) result in DofE. In our approach we predict an expectation and then verify its validity via the IS where both the expectation and assertion must be in A's model of B's beliefs and plan. Depending on A's own beliefs and plan, she can then respond accordingly, e.g., with "But these beans are fast cooking so it does not matter whether they're added first," if she disagrees with what she infers B's expectation to be but agrees with his assertion.

Recognising the defeated expectation in DofE is especially significant in TODs, since dialogue systems must be able to understand what expectations the user has in order to respond appropriately and collaboratively to the agents' goals, constraints and beliefs. For example, in the TRAINS example above, the procedure for NTOD presented in the previous section modelled B's expectation (in his private beliefs) as *be(ing) there at 11 > gives enough time to get to Bath*, which is highly specific and unlikely to be a static belief stored in B's private beliefs. It makes far more sense to assume the more generalised rule that *effect of plan so far > preconditions for next goal will hold*, but doing so requires mapping the situations described onto planning operators in the speaker's task plan.

5.2.4 Distinguishing plan-based "but" in TOD

Both Examples 5.3 and 5.4 are common examples of conflict in plan-based TOD signalled by "but" that does not involve causality at all. The utterances map onto planning operators in the task plan that are contingent upon one another, usually via satisfaction-precedence (s.p.) or dominance relations. We argue here that many of these planning relations convey expectations in a similar way to the defeasible rules in DofE in NTOD. Determining these expectations in TOD enables DPs engaged in accomplishing a task together to detect mismatches in each others' plans at an early stage and proceed with the task plan more smoothly.

Indirect SAs in NTOD are more difficult to resolve than the indirect action proposal in Example 5.3 since there is no plan to aid inference of the intentions behind the utterances. On the other hand, relations between planning operators proposed do not have to be considered in NTOD, though even NTOD cases require background and contextual information to be resolved if indirect SAs are involved.

Here we will focus on plan-based DofE in TOD and discuss how it can differ from NTOD cases like Example 5.1 by analysing what is being communicated. Example 5.5 above is hard to model because it involves DPs planning their future actions rather than carrying out their actions themselves as they discuss them. We consider simpler examples instead:

(5.7)
A: Add the vinegar to the sauce.
B1: (Yeah) But it's not tangy enough.
B2: (Yeah) But we forgot to add the mushrooms.

In the above example, B2 refers to steps in B's plan for making sauce which have not occurred. B1 refers to B's judgement on the result of adding the vinegar after tasting the sauce, where her perceived lack of tanginess triggers the "but". B1 involves causality, namely that adding vinegar makes things tangy, an expectation which is in her private beliefs and is violated by what she perceives to hold after the vinegar is added, so this is clearly a case of DofE that is very similar to the Greta Garbo example despite being a TOD.

There is no causality between A and B2, which simply involve satisfaction-precedence (s.p., $\prec$) between the given actions in the plan. Here B's plan contains the expectation *adding mushrooms* $\prec$ *adding vinegar*. This case cannot be viewed as involving a defeated belief that is static like B1; instead, B2 is a dynamic task-related expectation that will not be valid after the given stage in the plan has finished, and is clearly not a constant fact in B's beliefs the way that B1 is, which is a *generalisation* that holds regardless of context. Also, notice that B2 involves actions in the plan rather than effects like B1, and the only structural contingency that can occur between different actions in a plan are (1) s.p. which partially orders them in terms of when they operate in the plan and (2) dominance relations which specify which actions need to be achieved in order to accomplish higher-level actions in the plan. So distinguishing cases like B2 from B1 involves searching the plan for A and B and then determining where the conflict or defeated expectation arises.

In our approach, we adhere to searching B's plan for the task rather than some objective third-party observer's record of her actions. We claim that interpretation in dialogue is a subjective phenomenon, subject to the hearer based on their perception (as represented in the IS), rather than dependent on some objective model of the world.

This is where our approach diverges from Knott (1999a), who addresses monologue from the perspective of an independent observer.

Determining where these actions occur in B's plan and how they are related will indicate why B communicates contrast and what aspects of the plan are involved. In B2, B indicates a mismatch between what is been done already, and what should have been done according to her plan for making sauce. So it appears that these planning expectations in TOD are permissible only given a certain prior context and plan history. Whether this in turn should be viewed as a special case of DofE is uncertain. In either case however, such implications need to be reached from A by B in order to license the "but" in these examples, and will facilitate generation of an appropriate response to B's DofE.

5.2.5 Modelling DofE in TOD

In the previous section, we presented an procedure for DofE across speakers in dialogue, but this did not distinguish between TOD and NTOD. The treatment of plan-based "but" in TOD is often problematic in the NTOD procedure, since it does not involve the same sort of defeated expectations as in Example 5.1, but rather signals a difference in the speaker's plan for the task at hand.

The NTOD DofE procedure also does not address indirect SAs, which often occurs in TOD. Consider Example 5.6 above; here the indirect SA in A' makes the same proposal that A does, namely, that the speakers should add the mushrooms to the sauce. Applying this procedure to the assertion in A' results in the expectation *have mushrooms* $>$ *have added beans*, as opposed to the desired expectation *add mushrooms* $>$ *have added beans* which we would get if we interpreted the indirect SA as intended and substituted the planning goals it conveys instead of the facts communicated in the direct SA for the left-hand side (LHS) of the defeasible rule. So the interpretation of these defeasible expectations follows from the interpretation speakers understanding indirect SAs like A' above would normally make; i.e., they would take the indirect SA as the intended reading of the utterance. Here we also take into account indirect SA interpretation when formulating the defeasible rule defeated in cross-speaker DofE in TOD.

$$\left[\begin{array}{ll}
\text{PDU} & \left[\begin{array}{l}
\text{DH}\ \left[\begin{array}{l}\text{CA1: } action-directive(A, add(mushrooms, to(sauce)))\\ \text{CA1': } assert(A', assert(have(mushrooms)))\end{array}\right]\\
\text{TPH [TA1: } goal[add(mushrooms, to(sauce)), time = current, occurred = no]]
\end{array}\right]\\
\text{CDU} & \left[\begin{array}{l}
\text{TASK BEL}\ \left[\begin{array}{l}\text{TB1: } believes(B, [add_mushrooms > beans_added])\\ \text{TB2: } assumes(B, intention(add_beans))\end{array}\right]\\
\text{DH}\ \left[\begin{array}{l}\text{CA2: } assert(B, haven't_added_beans)\\ \text{CA3: } reject(B, CA1)\\ \text{CA4: } dofe(B, [add_mushrooms > beans_added])\end{array}\right]\\
\text{TPH [TA2: } action[add(beans), time = past, occurred = no]]\\
\text{COND}\ \left[\ accept(A, CA4) \rightarrow scp(A, dofe(B, [add_mushrooms > beans_added]))\ \right]
\end{array}\right]
\end{array}\right]$$

Figure 5.3: B: But we haven't added beans yet.

5.2.6 Representing Plan-Based DofE in the IS Framework

First we will discuss some features of the IS necessary to represent the information the procedure makes use of to resolve these cases. One significant modification to the NTOD DofE procedure involves looking up facts, which in the NTOD procedure were part of the speaker's private beliefs (and represented as such in the IS representation); here we distinguish between (static) private beliefs and task-related beliefs (TB field below) to emphasise the dynamic nature of these planning expectations. However as argued above, plan-based TOD cases of DofE involve checking facts in the speaker's planning history (Task Plan History, or TPH, in the IS), and expectations need to be checked by forward-chaining from the LHS (A's utterance) in B's plan to determine if this LHS precludes the negated right-hand side (RHS) from occurring.

We will make reference to IS fields in the procedure below to clarify how the interpretation will take place. The IS for the dialogue in Example 5.6 following B's utterance from A's perpsective of B's beliefs is shown in 5.2.6 with both A and A′ represented to show how they differ in the IS. The new abbreviated field is Task Plan History (TPH). Empty fields were omitted in CDU for brevity.

We assume that the procedure is triggered upon finding the "but" in B's utterance, and results in CA4 (Conversational Act 4) being added to the IS. This does not mean that CA4 is grounded; it only gets grounded as a *socially committed proposition* (or *scp*,

following Matheson *et al.* (2000)) after A accepts it, since (as Matheson *et al.* (2000) argue) A has a social obligation (in order to be cooperative) to respond to B's utterance. Prior to distinguishing DofE, the indirect SAs, which in TODs are often planning proposals, need to be resolved, so e.g., A′ needs to be resolved to communicate the same SA that A communicates (i.e., CA1 and CA1′ in the IS above). This occurs via a Conversational Act (CA) interpretation process that utilises a planner which gets called after each new utterance and can recognise the implicit proposal in A's assertion (when passed the logical form of the utterance) and detect evidence contrary to this proposal in B's plan. This means that the IS is kept updated with current task actions communicated via scripts that communicate with the planner and request forward-chaining or plan-recognition with various inputs. The interpreter maintains communication with the dialogue manager in order to keep the IS updated with what is happening in terms of the agents' salient planning actions after each turn in the dialogue which are stored in TPH.

Mapping utterances into planning operators and SAs is feasible in restricted TOD domains and has been implemented in other systems (e.g., the BEE tutorial dialogue project[6], the COLLAGEN project (Lesh *et al.* (1999)) and work by Lambert and Carberry (1991)). Notice in the IS above that A needs to assume that B has the intention to add beans. The procedure below operates on the semantic representations of utterances recorded in the IS, post CA interpretion. We gloss over construction of semantic representations for utterances here, so interpret "utterance" in the procedure as referring to this (quasi) logical (and highly stripped down) representation, e.g., $have(mushrooms)$ for A′.

5.2.7 DofE in TOD Interpretation Procedure

If the utterance in CDU contains a "but", test to see if DofE holds between CDU and the utterance in PDU: (For now we assume that these are just the two adjacent cross-speaker turns sandwiching the "but".)

1. Interpret the propositions in CDU.DH via the planner to determine what task operations/actions/plans they communicate, update CDU.TPH with this new task

[6] See www.cogsci.ed.ac.uk/~ jmoore/tutoring/index.html.

action, and then negate[7] it. *So "But we haven't added beans yet" becomes action[add(beans),time=past, occurred=yes]*
Add to CDU.DH a CA for B's assertion that beans have not been added.

2. Find the task action interpretation of the previous turn in PDU.TPH (task plan history) and make this the LHS of the rule. *So here, LHS = goal[add(mushrooms,to(sauce)),time=current, occurred=no]*

3. We now have the rule ($LHS > RHS$), and we pass it back to the update module. *Rule:* $goal[add(mushrooms, to(sauce)), time = current, occurred = no] > action[add(beans), time = past, occurred = yes)]$.

4. The update module adds this rule to A's representation of B's task-related beliefs, CDU.TB (after hearing the "but" utterance). We also add to CDU.TB the assumption that B has the intention to add beans, and add the DofE relation to CDU.DH as one of the CAs along with B's rejection of A's action-directive. Before A's turn arrives, this gets added to CDU.

5. We check that the rule is in A's plan only when A's turn arrives in order to generate the appropriate response.[8]

Interpreting the utterances as the task actions they communicate allows us to interpret DofE in cases where task goals are communicated implicitly, i.e., via indirect SAs, as in Example 5.6A′. The NTOD DofE procedure could only address direct SAs and would not have been able to address indirect SAs at all. This TOD DofE procedure enables interpretation of DofE involving indirect SAs in TOD, where a planner and CA interpreter are available and the domain is restricted.

Notice also that the defeasible rule interpreted (i.e., $goal[add(mushrooms, to(sauce)), time = current, occurred = no] > action[add(beans), time = past, occurred = yes)]$) assumes that the speaker planned on adding beans if mushrooms

[7] Negation might be too strong in some cases, and we might want to use an alternative set semantics instead. However negation does tend to represent the situation quite well in TOD, where situations often involve actions either happening or not, while NTOD often involves non-binary situations where alternatives better represent the process involved.

[8] This is elaborated in Chapter 9.

were also being added and simply expresses action s.p.. The task action interpreter will keep track of planning operator pointers in the task plan corresponding with the dialogue, so that referencing the plan is simple. Checking whether the expectation/rule is in A's plan in step 6 involves simply determining whether it is the case that the LHS of the rule s.p. the RHS, possibly recursively if necessary. Assuming that the speaker intended on adding beans is implicit in B's utterance, since if he had not intended on adding them, he would have no need to point out that they have not been added. A can lack this assumption about adding beans but still have the rule in her planning expectations, which licenses her response of "oh yeah, I forgot" if that is the case. If she does not have the rule/ordering of adding ingredients then she can adopt the idea of adding beans or indicate that she had not planned on adding them, as will be discussed in more detail in Chapter 9.

The inferred expectation is stored in B's task-related beliefs rather than his private beliefs to capture the distinction between plan-based DofE in TOD and expectation-based DofE in either TOD or NTOD, since expectations in NTOD like "beautiful people usually marry" and "adding vinegar makes things tangy" are static and remain in private beliefs, unlike plan-based expectations like "adding mushrooms means beans have (usually) already been added", which have a shorter "life-span", and are valid only in the context of following this particular recipe for soup, and perhaps only at a few particular stages in this plan. Plan-based beliefs are dynamic, and their validity is tied to planning context in a way that expectation-based beliefs are not, since one can imagine the belief about beautiful people marrying lingering in the speaker's private beliefs indefinitely (though of course only salient given dialogue context), while the belief about adding beans before mushrooms might be invalid given a different recipe or even different stage in the plan (e.g., consider the situation where an action is repeated, but with differing preconditions, in a plan).

5.2.8 Implications for Monologue

Differences in both speakers' beliefs and plans are phenomena unique to dialogue rather than monologue. While our approach relies on the IS model of dialogue which represents different speakers' beliefs, it should be possible to extend this treatment to

monologue within the same IS framework, as was done for instructional text generation in Zaenen *et al.* (2000). For generation, knowledge of a speaker's beliefs (and plans, in task-oriented texts), as well as a representation that includes what is already been said and captures these distinctions, as in the IS approach, enables generation that expresses relations like DofE and concession by representing connections between new information and old expectations. E.g., a speaker may have the expectation that adding vinegar makes things taste tangy, and if she then finds that the soup she is making is not tangy despite adding vinegar, she might say (e.g., TV chef instructors) "I added vinegar to this sauce, but it still didn't taste tangy." (If she were on TV, she would have to be knowledgeable, so she would probably then explain why this is the case, e.g., "This [DofE] is because the cream absorbs the acidity of the sauce".)

Signalling DofE with "but" given a defeasible expectation and having just discovered a contradicting fact is useful for text generation systems that need to respond dynamically to situations as they arise. Our extension of interpreting CAs w.r.t. task plans enables generation of indirectly expressed DofEs as well, e.g., "I thought some vinegar might be nice, but it still didn't taste tangy". The procedure presented provides the basis for modelling DofE, both in TOD and task-oriented monologue (TOM); the main modifications necessary to address monologue would involve searching the Monologue and Plan Histories for inconsistencies between defeasible expectations and facts (current or forthcoming, e.g., the effect of some current action) before generating a new utterance, and if an inconsistency is detected, then DofE should be generated. Interpretation of DofE and responding to it is unnecessary in monologue, and much of the cross-speaker aspects of the procedure can be eliminated.

5.3 Incorporating Underlying Relations

In this section we will explore the nature of the expectations denied in DofE further; i.e., we will investigate how information about the relation licensing the underlying expectation/rule can be used in our DofE interpretation procedures. We will assume that the methodology for extracting these underlying expectations proposed in section 5.3.2 is in place, and that we have access to the resulting feature-value relation defi-

nitions for the expectations, which can be represented in the IS, as will be addressed below. Modelling these feature-value descriptions of the relation beneath the expectation in DofE can improve generation by enabling more precise responses that address these underlying relations. Example 5.8 below illustrates the idea that underlying expectations denied in DofE often involve the denial of noncausal relations across turns in dialogue, contrary to the usual assumption that DofE denies causal expectations (Lagerwerf (1998)):

(5.8) A: Greta had a child in ’43.
B: But she married in ’47.

Here the relation is a temporal one between getting married and having children. In TOD we saw satisfaction-precedence (s.p.) relations across turns triggering expectations. The difference between the s.p. and temporal ordering cases like Example 5.8 is that in NTOD there is no notion of the agency of speakers/agents which is inherent in TOD. I.e., the accomplishment aspect of s.p. is novel to planning, where goals are posted and accomplished, and there is a sense of agency. Temporal ordering relates actions, events, states, effects, etc, with no notion of agency involved.

Temporal ordering is not the only relation that can underlie denied expectations across turns in dialogue. Consider for example a dialogue version of the RST (Mann and Thompson, 1988) example of the *evidence* relation, where the evidence relation is denied:

(5.9) A: The program really works.
B: But I didn’t get a result which agreed with my hand calculations.

Normally, getting a result from hand calculations which agrees with what a program predicts is evidence (though incomplete) for the claim that the program works (A’s assertion), although this example could also be viewed as involving causality, or restatement, since it is a condition for a working program to produce correct results, which is denied here, undermining the credibility of A’s assertion. Evidence can also be viewed as an interpretation of expected/desired effects in a plan-based situation, though here both A and B express states rather than actions. Here we see denial of a piece of evidence for A’s turn, i.e., this example can be viewed as a DofE where the expectation is evidential, since B’s turn if inverted to its positive form, i.e., “I get a

result which agrees with my hand calculations", contains evidence for the claim made in A.

Inverting "but" turns

When we talk about inverting B's turn to reveal its positive form in these DofE cases, generally we mean removing the "but" and removing negation from the main clause verb phrase, so "but I didn't get a result" becomes "I get a result". Inverting Example 5.8 would result in "She didn't marry in '47". Notice that here, the expectation being denied, "having a child in '43 $>$ not marrying in '47" does not make much sense, because it does not get at the underlying temporal ordering of marriage preceding having children that is being denied. We will follow this simple model of inverting "but" turns to get at the original denied expectation and discuss how relations like these are modelled in the next few sections.

5.3.1 RST Relations: Possible Candidates?

So one systematic approach that might be utilized to investigate the range of underlying relations in these DofE cases is to use a set of rhetorical relations like those in RST (Mann and Thompson, 1988). We must bear in mind that RST was never designed to address multi-speaker dialogue let alone spoken rather than written text. However we consider it briefly as it is one of the most widely used theories addressing discourse relations in applications. Furthermore, there are no theories of discourse relations that specifically address cross-speaker relations in dialogue aside from work on question-answering and correction, e.g., Asher and Lascarides (1998b), Asher (1998) and Poesio and Traum (1998). The more coarse-grained relations in SDRT (Lascarides and Asher, 2002) do not help interpretation of the relations underlying these denied expectations in DofE.

One difficulty which arises with RST is how to frame *presentational* relations, like the evidence relation in Example 5.9 above, across turns in dialogue. This is because presentational relations (according to RST) are ones communicated with the intention that they produce an effect (or increase some inclination) in the hearer, and are distinguished from *subject-matter* relations which have only the intended effect that they are recognised as holding between the related clauses. For concession, presentational re-

lations which relate material across speaker turns requires that the speaker of the cued clause (B) accepts the previous turn at least partly; though minimally, B must simply interpret A's turn in order to refer to it as an argument in the concession relation he is communicating. In terms of intentions, this means that B must intend to relate A's turn to his own turn to create some inclination in the hearer (i.e., A) of the cued turn (B's DofE).

In any case, a clear distinction must be made between which aspects of the relation are intentional and which are informational. Moore and Pollack (1992) argue that often the intentions of the speaker can be inferred from the informational content and vice versa, since the reasoning processes between informational and intentional content are related. Also there is no isomorphic mapping between intentional and informational relations, so inferring one from the other requires additional information. Moore and Pollack also illustrate cases in which the same discourse can postulate entirely different structures of relations with different nuclei, etc. In short, RST relations are quite underspecified. The example above of the presentational relation evidence (adapted from the RST text for dialogue) does not work reliably across turns when the evidence relation is not being denied (which is a further piece of evidence that the presupposed defeasible expectation is cued by "but"):

(5.10) A: The program really works.
B: Yup. Every time I get a result which agrees with my hand calculations.

A brief discussion on nuclearity

Notice in the example above that B needs to explicitly indicate agreement with A first, and then his "Every time ..." serves to explain why he agrees, rather than provide direct evidence for A's assertion. That is, if an RST tree were drawn for the two turns (ignoring cross-turn difficulties by assuming a cross-turn agreement relation), then it seems most reasonable to argue that the "Every time...calculations" sentence is the satellite of the *evidence* relation it shares with the "Yup" (the nucleus). This unit (comprised of B's whole turn) can then be seen to be involved in an *agreement* relation with A's turn. Even if we argue for a different RST structure, the ambiguity in interpreting a unique structure for this example lends weight to the claim that RST is quite underspecified. If RST were modified to address two speakers' intentions, then

the notion of nuclearity across speakers would need major revision, since assuming equal initiative in the dialogue, one speaker is unlikely to be subservient (i.e., in a satellite relationship) to another speaker's intentions.[9]

5.3.2 Distinguishing Underlying Expectations

Given the difficulty of using RST to distinguish the relations underlying denied expectations across speaker turns, we present a novel methodology for distinguishing features involved in these relations using linguistic substitution tests involving the cue phrase taxonomy presented in Knott's thesis (Knott, 1996). Knott presents a taxonomy of cue phrases distinguished as feature-theoretic constructs rather than markers of one or more of a set of rhetorical relations as postulated in RST. Rather than finding data to describe a conceptualised theory of discourse relations, he uses data containing cue phrases to drive the creation of his taxonomy of cue phrases, which reveals psycholinguistic features involved in conveying or interpreting meaning, i.e., the data drives his theory of linguistic production. We enquire into the nature of the relations underlying these denied cross-turn expectations using the following methodology:

1. Take original "but" example and determine expectation being denied via DofE procedure

2. Invert example so that the consequent of the expectation is asserted rather than denied in B's turn, (i.e., omitting the "but"). (So the dialogue conveys that the expectation in Step 1 succeeds.)

3. Determine what sort of expectation this inverted pair of turns seems closest to, given Knott's taxonomy. Determine whether the cues conveying this relation are substitutable in this inverted dialogue:

   (a) Determine what the nature of the expectation involves by asking how the antecedent and consequent are related; if unsure, test all the high-level categories in the taxonomy and see which ones work by substitution tests

[9] However in situations in which one of the speakers is markedly dominant (e.g., situations with little mixed initiative), or in which the speakers are both involved in a narrative and each turn continues the preceding story with virtually no disagreement, we might see different behavior.

involving cues belonging to those categories. One way to do this might involve examining whether the relationship between antecedent and consequent involves causes or results, whether it appears to be uni- or bidirectional, whether one or the other is a part or superset of the other, etc. Then determine whether the category chosen captures the nature of the expectation.

(b) If so,

i. test whether hyponyms[10] of these high-level cues work in the inverted dialogue. The most specific hyponyms that work indicate the maximally specific set of features that pertain to the relation underlying the expectation.

ii. Now confirm that these cues that work in the inverted dialogue do not work in the original (denied) dialogue (i.e., if they express the underlying relation, they should not work when the relation is being denied). Those cues that work in the inverted example but not in the denied (original) dialogue are indicative of the nature of the relation underlying the expectation that is denied in DofE.

iii. Look up the feature-value definitions for the maximally specific cues that work in this inverted (not denied) case.
Take the intersection of feature-values for the maximally-specific cues in the inverted example and compare this with the intersection of feature-values of hyponyms of "but" that deny the same expectation (i.e., in the original dialogue) to reveal which feature-values are denied/inverted in the denied case. This tells us precisely which aspects of the expectation are being denied!

(c) If not, i.e., if Knott's taxonomy does not provide a category that works for the inverted dialogue,

[10]Hyponyms in Knott's thesis inherit the features of their parent (higher-level dominating) cues in the directed acyclic graph structure of the taxonomy. So if a cue is substitutable, then all of its hypernyms (higher-level parents) should also be substitutable in that case. The higher-level cues are far less specifi c and so convey less precise information, but they should still be substitutable whenever their hyponyms work.

i. then check what the original denied expectation would involve and check whether any of Knott's categories fit by testing which cues are substitutable in the original example; a good place to start is with hyponyms of "but".

ii. For cues that work, check these cues' hyponyms to determine the maximally specific set of features that apply to the relation between turns. Note that this only specifies the relation underlying the denied expectation and does not shed light on the original (not denied) expectation.

iii. If no cues besides "but" work in the original dialogue, then "but" must be the maximally specific cue that works, and we can not determine more precisely the nature of the denied expectation, so assume that the turns are related by simple contingency/co-occurrence for now.

An example

Taking Example 5.1 and applying this methodology, we get for the inverted dialogue:

(5.11) A: GG was beautiful.
B: (?Yes,) <Cue> She married.

If we insert the "yes" here, we assume B accepts A's assertion. For the "but" cases we assumed two possibilities: (1) B implicitly accepts A's assertion but introduces additional information which is somehow odd given A's assertion, and (2) B understands but disagrees with A's assertion and asserts something which is odd, thereby communicating a reason why A's assertion might not be true. In either case, assuming that B understands A does not seem too strong in the inverted dialogue above.

So if we analyze the expectation involved above by investigating which sorts of cues are substitutable in this dialogue, we will hopefully get a sense of what sort of relation the expectation conveys. To determine whether the two properties (being beautiful and being married) are both inferrable from each other (bidirectionality), we ask if the marriage precedes being beautiful or vice versa; since either seems plausible to a third-party observer (the author), it seems reasonable to assume that a bidirectional contingency is involved, so we can tentatively rule out unidirectional relations like cause or result. This just leaves "additional" information relations, which seem to be permissible and is the closest we can get via Knott's taxonomy to the idea of

contingency or increased likelihood of two properties holding in conjunction. So testing these additional information cues, and (indicating possible acceptance with "?", acceptance with no marker, and unacceptable with "*") we have:

(5.12) A: GG was beautiful.
B: (?Yes,) <in fact/indeed (?and)/?as a matter of fact/?actually/ even/*on the contrary> she married.

Some of these additional information cues work given the assumption that B has the expectation that beautiful people usually marry. So it seems reasonable to allow that an additional information relation holds between A and B's utterances. But this does not get at what sort of information constitutes additional information or how the additional information is related to the prior information beyond a loose assumption of new information on a common topic.

A question that arises with this methodology is whether the expectation that the contrary holds arises because it is triggered by the "but" alone, in which case, our method of removing the "but" to get at the type of relation conveyed is probably wrong. We argue that given this sort of contingency expectation, it should be no less likely from B's perspective when generating a response, to indicate a contingent expectation which fails to hold (DofE) than to indicate a contingent expectation which holds. In both cases B communicates this expectation, though from an interpretation perspective (of B's turn), without the "but" we need the presence of an additional information cue to infer the expectation. So in either case, the expectation is still present regardless of whether it is denied or asserted in B's turn. Presumably the speaker has in her private beliefs a bunch of contingent expectations that tend to correlate with the property or event described in A's turn. They all become salient when she interprets A's utterance. If she knows that one of these correlated properties or events does not hold, she expresses DofE that this property does not hold (e.g., "but she never married"). If she knows that some of these correlated properties or events holds, she might offer this as additional information as in the example above ("yes, in fact she married").

So this means that the expectation underlying this dialogue introduces additional information. But this seems unsatisfying, because it does not capture the fact that B seems to correlate these two features and infer a bidirectional contingency from one to the other. So how do we investigate this further? Possibly by testing cues which

express the opposite relationship to contingency and seeing whether these cues are all nonsubstitutable. So the opposite of contingency and correlation is mutual exclusivity or at least properties that tend not to correlate or occur together, rather than vice versa. The idea here is that in isolating first the lowest level cues that work in the original dialogue, we have a set of relations (as feature-value constructs) that represent the underlying meaning; by inverting the dialogue and finding the most specific cues that work here, we have the "opposite" relation's feature-value bundle. By comparing the features in common which have opposite values (e.g, positive polarity in the inverted dialogue and negative polarity in the inverted case), we get at the relations which are being denied in these cases; i.e., the features that have different values in the original and inverted cases together represent the underlying relation communicated in DofE. In the inverted dialogue below, negative polarity cues all fail, whereas positive polarity cues failed in the original case.

(5.13) A: GG was beautiful.
B: (?Yes,) <*instead/*rather/*but/*however/*even so/*in spite of this/ *although/*while/*even though/*whereas/*on the other hand/*having said that/*all the same/*nevertheless/*despite this/*though/*alternatively> she married.

The asserted (inverted) expectation does not seem to license negative polarity cues. Now we revert to the original minidialogue and make sure that the unacceptable negative polarity cues are acceptable here:

(5.14) A: GG was beautiful.
B: <*instead/*rather/but/however/even so/in spite of this/ ?although/*while/?even though/*whereas/?on the other hand/?having said that/all the same/nevertheless/despite this/though/*alternatively/then again/*unless/*otherwise> she never married.

This shows "but" and several hyponyms, all of which preserve the meaning here, and many of which are acceptable. So we might want to investigate the cues which are not acceptable to find out which relations specifically are not being communicated here. "Instead" is contingently substitutable with "but" and is not a hyponym, and the same holds for "rather"; "while", "whereas", "alternatively" and "on the other hand" are hyponyms but are all not substitutable for "but".

Distinguishing features

Taking some of the maximally specific (lowest level) cues that work in the denied (original) example, we see that for "despite this", Knott's feature-value bundle has pragmatic source of coherence, anchor is cause-driven, pattern of instantiation is bilateral, focus of polarity is count, polarity is negative, it is not presuppositional, involves actual modal status and has causal rule type (from Knott's taxonomy); for "then again" the only contradictory features are pattern of instantiation (which is unilateral here) and rule type (which is inductive). So leaving these two features unspecified (i.e., taking the intersection of feature-values for the maximally specific cues), the remaining features that are specified are the most specific set of features that describe the underlying denied relation.

So to most specifically describe what is being denied in the expectation, we determine that source of coherence is defined as pragmatic and anchor is defined as cause-driven (where both were undefined in the inverted dialogue). Focus of polarity switches value to being counterpart-based and polarity switches to being negative, and these new values are (we argue) the best way to characterise what exactly is being denied in the DofE. Knott provides clear definitions of the features in his thesis that allows categorisation of cue-phrases.

We can then investigate precisely which feature-values are being inverted in the denied case by comparing the intersection of feature-value bundles for the maximally specific cues of both the asserted (inverted) and DofE (original) dialogues. The values of features that (1) differ and (2) are specified in one case and not in the other describe precisely how the expectation being denied in DofE is denied. So between Examples 5.13 and 5.14, given the intersection of maximally specific cues in both asserted and denied dialogue, we see that the values being denied are polarity and focus of polarity, as illustrated in Table 5.1.

Generation Benefits

Now the hearer of the DofE can interpret these features being denied and determine his own response based on how these feature-value pairs are defined in his own model of beliefs. So here, if he disagrees with the polarity assignment to the asserted and denied cases (i.e., he views the inverted dialogue in Example 5.13 as characterised by negative

Table 5.1: Feature-Values of Relation Underlying Expectation in Examples 5.13 and 5.14

Features	Asserted *indeed* *even*	Denied *despite this* *then again*
Polarity	Positive	Negative
Source of Coherence	–	Pragmatic
Anchor	–	Cause-driven
Focus of Polarity	Anchor	Counterpart
Presuppositionality	Non-presupposed	Non-presupposed
Modal Status	Actual	Actual

polarity with the "but" making sense there, and the original dialogue (Example 5.14) as positive with "indeed" replacing "but"), he might address this and say "but not marrying is common among beautiful people. In fact, it would be odd if as a beautiful woman she married." Given these tests (backed by the reasoning discussed here) we can see that the inverted expectation does not just introduce "additional" information, but also introduces bidirectional contingency between this new information and the preceding turn involving positive anchor-based polarity.

While it might seem that Knott's features are too fine-grained to predict information that is useful for a dialogue system, we claim that these fine-grained features are precisely what is needed to determine the different ways in which the denied expectation differs in its denied versus asserted forms to more closely target speakers' differences in beliefs.

What this mechanism will require is a means of interpreting pairs of propositions in Knott's taxonomy in order to determine precisely which feature-values are involved. Currently this is determined via human judgment, i.e., linguistic substitution tests, and so the procedure cannot be automated. So the hearer of the DofE (A) needs to interpret B's DofE with respect to her own preceding turn to determine the feature-values being denied (in her own model of the world). Once she expresses her opinions about what she perceives B to be denying and the relationship she assumes B uses to

link the propositions, she responds accordingly, giving B the opportunity to correct any misunderstandings, and thereby align their beliefs (or detect discrepancies) much more directly.

Representing broader categories

Besides defining cue-phrases and locating them in the taxonomy based on specific feature-value bindings, Knott places cue-phrases in several broad groups which do not necessarily share common characteristics with respect to feature-value bindings. These broader categories are grouped as follows: sequence, cause, result, restatement, temporal relations, negative polarity, additional information, hypothetical relations, similarity and digression relations. These groups contain cue-phrases which can exclusively or nonexclusively belong to one or more categories. As a rough guide to the nature of cue-phrases, including this information can be quite useful. So we can gain more general information about the nature of the relation underlying the expectation by considering the category that the maximally specific cues belong to. So, e.g., for Example 5.8 (i.e., having a child in '43 but marrying in '47), we can represent that the DofE hinges on temporal ordering, as will be seen in the next section.

5.3.3 Modelling Issues

Although the methodology presented in the previous subsection requires human judgment to assess the results of the substitution tests, it is a first step towards distinguishing underlying relations in cases in which the material in the two turns across the "but" directly leads to a denied defeasible expectation. So while we do not present procedures that automatically determine the nature of the relation underlying the denied expectation in DofE cases, we address how this information might be modelled in the PTT model of dialogue by adapting the IS representation (Matheson *et al.*, 2000) to reflect this new information.

5.3.4 Updating Features

As we have seen, Knott's feature-theoretic view of cue phrases defines relations between clauses in discourse via a bundle of feature-value pairs (Knott (1996)). We argue

that Knott's data-driven definitions for cue-phrases as markers of relational constructs defined via a bundle of feature-values for a set of features motivated from the data is preferable to assigning hypothetical relations that are not generated from real data. However, while Knott's cue-phrase definitions are very useful due to their specificity, we run into the problem of mapping these abstract feature-value pairs onto expectations in terms of real dialogue. I.e., how can we interpret what is being communicated when handed a feature-value bundle, and how does this help us predict what sort of response to generate, depending on the other dialogue participant's beliefs w.r.t. this relation?

5.3.4.1 Utilising Knott's Feature Definitions

Knott argues that his data-driven definition of relations is compatible with the view that relations are planning operators with preconditions and effects, where the relations' preconditions are defined via the speaker's intentions and applicability conditions specified for what the speaker wants to convey, and the effects are simply the intended effects the conveyed relation has on the hearer. More practically speaking, the features are defined in terms of variable bindings and relationships which describe the relations concisely. For example, the *polarity* feature describes whether the defeasible rule $P > Q$ holds, based on whether A=P and C=Q (positive) or A=P and C is inconsistent with Q (negative), where A and C are the propositional contents of the two respective related clauses. Some relations allow one to make broader generalisations, e.g., inductive rules, which allow one to generalise that a property holds for all members of a given class if it holds for some. To address how polarity might be determined in a dialogue situation, if a speaker believes $P > Q$, then this is in her Private Beliefs field in the IS. If her turn is mapped onto Q, and the prior turn is mapped successfully onto P by matching first-order logic representations of the material in the two turns, then if her turn maps onto Q, we can assume positive polarity; if her turn maps onto a negated Q, then we assume negative polarity.

While mapping speakers' turns onto the variables which define Knott's features might be difficult, we can still utilise Knott's feature-value definitions for cue-phrases assuming we have found the most specific cue-phrase which works in the given di-

alogue following the methodology in the preceding chapter. We can automate some of the feature assignment to update the IS by maintaining an exhaustive (i.e. complete) static table of cue-phrase definitions[11], despite the need for human judgement to perform substitution tests, which cannot be automated.

However, once the most specific cue-phrases that work in the inverted and denied expectations are determined, we can automatically assign feature-value-pair bundles to these dialogues which describe the underlying relation being denied. Then comparing the feature-values for the maximally specific cues for both the asserted and denied cases (as we saw in the previous section), we can determine precisely which features are being denied in a given DofE, and the IS can be updated with this information, so that in the next turn of the dialogue, the speaker can compare these feature-value assignments to his own (in his private beliefs) and respond accordingly with a highly specific response to the DofE which targets precisely where he disagrees or agrees. We will need to either find a way to automate the parsing of pairs of beliefs in the IS to associate feature-value bundles with speakers' beliefs, or require a human annotator at this stage.

5.3.5 Information State Modification

Recall that in the "but" triggered cross-speaker DofE in our approach, the IS was updated with a CA containing the DofE, speaker whose expectation was denied, and the denied expectation, e.g., CA: $dofe(B,[beautiful(X) > married(X)])$.

We propose, given information about the nature of the underlying relation via the feature-value differences involved in the DofE as well as broader information about the category(ies) to which a cue-phrase belongs in Knott's taxonomy (Knott (1996)), to include this in the $dofe$ CA as follows for Example 5.1:

dofe(B,$[$*additional_information(beautiful*(X)*,married*$(X))]$*,*$[$*polarity,focus_of_polarity*$]$*)*

since we know that the expectation in Example 5.1 simply describes a perceived contingency, which, after applying the methodology proposed in the last chapter, translates to Knott's "additional information" relation holding between the clauses, where polarity

[11] Knott provides a partial table of cue-phrase definitions like this in Appendix D.1 of his thesis, (Knott, 1996).

and focus of polarity feature-values are denied. For Example 5.8, the new representation would be $dofe(B,[temporal\_ordering(marrying(X,t),having\_children(X,t');t < t')],[])$, since here applying the methodology indicates that Knott's "temporal ordering" relation best describes the underlying expectation. It does not make sense to indicate that marrying defeasibly implicates having children, since this is not what the expectation involves (at least not explicitly). This expectation reflects B's relative temporal ordering of the two events, and does not imply simple correlation as in Example 5.1, where B expects beautiful people to marry.

Likewise, for TOD we will follow the procedure and update the IS with the predicted planning relationships if this is how the turns are related, so Example 5.7B2 would have $dofe(B,[s.p.(add\_mushrooms,add\_vinegar)],[])$, where the last field would include features being denied.

5.3.6 TOD vs NTOD

In cases in which specific features are the precise source of the DofE, e.g., in Example 5.1, if the hearer of the DofE can recognise that the wrong polarity is being attributed to his utterance, he (A) might indicate this misassumption by saying "but not marrying is common among beautiful people". However in TOD, these sorts of misunderstandings can result in much confusion. E.g., consider Example 5.7B2. Following our NTOD approach, we would predict that B has the expectation that $adding\ vinegar > haven't\ forgot\ to\ add\ mushrooms$. I.e., the procedure recognises contingency relationships, since it simply negates the situation described in B's turn to produce the consequent of the inferred expectation. This means that A has failed to recognise that B has an expectation that adding mushrooms s.p. adding vinegar. We address these s.p. situations in TOD by searching B's task-plan, and predict for Example 5.7B2 $goal[add(vinegar),time = current,occurred = no] > goal[add(mushrooms),time = past,occurred = yes)]$, which captures the expectation about temporal ordering of events implicitly but relies on the planner to determine what sorts of task planning operators they communicate. So while the procedure in the previous section works for TOD, it cannot handle NTOD DofE like the example above. And yet the denied expectation in Example 5.8 above involves temporal ordering between events discussed

in the dialogue much like the TOD Example 5.7B2.

Intuitively these cases seem very similar in terms of the sort of expectation denied, but in practice, the NTOD cases are much harder to resolve due to the lack of a small domain and planner which can map the turns onto planning operators and determine both whether they have occurred or not and also whether they have a particular ordering in the speaker's task plan (or beliefs). What we claim here is that the two cases are examples of the same phenomenon, i.e., that there are relations underlying these denied expectations in both TOD and NTOD. While we do not present a machinery for determining these underlying relations in NTOD (since we must rely on human judges of substitution acceptability), we have shown here how, once determined, they might update the IS in the PTT model of dialogue and be used to generate more appropriate responses depending on the hearer's own beliefs.

5.4 Summary

This chapter modelled DofE for both TOD and NTOD in the PTT IS model of dialogue. The last section addressed where information about relations underlying denied expectations can be incorporated into the model, assuming the methodology presented in section 5.3.2 has been applied.

Despite the apparent dissimilarity between TOD and NTOD procedures for distinguishing and modelling DofE, similarities for DofE between the differences in discourse type (TOD vs. NTOD) have emerged; e.g., we have seen how temporal ordering of events is somewhat similar in both dialogue types. Both TOD and NTOD involve the same sort of reasoning to interpret the denied expectation, but TOD needs to determine these denied expectations via task-plans while NTOD involves searching beliefs. Also in both cases, it is often the case that a more specific relation than simple defeasible implication is possible, and hopefully it has emerged that representing these relations more specifically is better than leaving the turns related only by unidirectional contingency. The central idea of the methodology we presented to get at this underlying relation involves inverting DofE dialogues to get at the features which values switch (e.g., positive to negative polarity) between the inverted and original dialogues. The set

of features which change and their corresponding values for the original DofE provide a data-driven mechanism for representing the underlying relation.

Although we focus here on A interpreting B's DofE, presumably similar reasoning is involved in the generation of DofE. We discuss generating "but" triggered cross-turn relations more in the last chapter, as well as how one might deliberate over a response to cross-turn DofE.

Chapter 6

Concession

In this chapter we will consider a rhetorical device typically used in monologue as an argumentative tool whereby speakers agree to or accept part of an opposing argument while arguing for an alternative perspective which is stronger (i.e., more convincing) than what is agreed to. For example, "[We'll be really tired if we don't return until Sunday], but [the extra time in Israel will be worth it]."[1] Here the speaker acknowledges a downside for staying longer on holiday, but concludes that this shortcoming is worth the longer holiday. RST[2] (Mann and Thompson, 1988) defines concession in monologue texts as having the following conditions:

1. The speaker thinks the situation of the Nucleus is true, valid, good or preferable.
2. The speaker is not saying that the Satellite is not true.
3. The speaker acknowledges a potential or apparent incompatibility between the Satellite and Nucleus.
4. The speaker thinks that the Satellite and Nucleus are compatible.
5. If the hearer realises that the situations of the Satellite and Nucleus are compatible, the hearer's positive view of the situation in the Nucleus will be increased.

RST does not specify compatibility conditions, and it is not clear whether the potential or apparent incompatibility in step three occurs either at a future point in time or given certain other additional conditions, leaving steps three and four asserting possibly contradictory conditions. Concession is a presentational relation in RST, that is,

[1] This example was taken from Stent and Allen (2000).

[2] The concessive cues presented in RST include "although", "though", "however", "even so" and "nonetheless". They probably exclude "but" since it is more general and can be used to cue other relations as well, e.g., semantic opposition.

one which is intended to produce some effect in the beliefs of the hearer with respect to the argument being made.

In a different approach, Lagerwerf (1998) and Kruijff-Korbayová and Webber (2000) argue that in an example like "Although John plays basketball, he doesn't play football", the conceded clause ("John plays basketball") leads to the inference of the claim "John is a sporty person", which is then denied by the second clause in the sentence, which proposes an antecedent (i.e., "he doesn't play football") which counters the inferred consequent of the subordinate clause. In the case of "but", which is a conjunction rather than a subordinating adverbial, generally the clause following medial "but" contains the clause argued for. So the first clause in a sentence with medial "but" contains an argument which allows the inference of a claim which is then denied by the second clause.

In this thesis, we consider turn-initial "but" (which corresponds loosely with medial "but" in monologue), and in this chapter, we will focus on modelling the concessive analogue of "but" in dialogue. Given the intentional aspects of the presentational relation (following RST terminology) concession[3] one might wonder how a phenomenon involving one speaker's intention to present a stronger alternative argument by conceding an opposing viewpoint can be stretched across turns in dialogue, where different speakers utter the conceded and counter clauses, as in the following example:

(6.1) A: John plays basketball.
B: (Yes) But he doesn't play football.

Here the same example occurs in dialogue, and we view this as a form of cross-speaker concession, since B agrees (explicitly or implicitly) with A's claim despite presenting an alternative perspective with respect to the inferred claim that John is sporty. What is happening in this case is that B agrees with A's assertion but disagrees with the inferred argument that John is sporty by presenting a counterargument (since not playing football might lead one to the conclusion that John is not an overall sportsman).

Recall that Denial of Expectation (DofE) as discussed in the last chapter would involve B saying "But he's not a sporty guy", directly denying the inferred conse-

[3] See Moore and Pollack (1992) for discussion on intentional and informational aspects of RST relations.

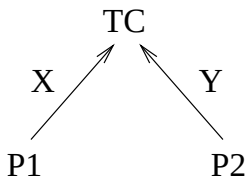

Figure 6.1: Interpretation schema for concession across turns

quent. Concession can be distinguished from DofE since it does not explicitly deny the consequent as in DofE, but rather presents an alternative antecedent that favours the inference of the negation of the claim inferred from the claim in the previous turn. In other words, "John plays basketball" allows inference of the claim "John is sporty" which B's assertion argues against, i.e., John not playing football leads to the inference that John is not a sporty guy overall.

In this chapter we will examine different situations involving speakers (1) agreeing to or accepting part of an argument inferred from the preceding turn and then (2) presenting an alternative argument w.r.t. the inferred claim. We will see that concession does not have to involve directly opposing the claim inferred from the other speaker's turn (as in the example above), and can often involve the presentation of alternative perspectives w.r.t. the claim. That is, we will view concession as involving the basic scheme illustrated in Figure 6.1, where the concessive propositions in the two turns ($P1$ and $P2$ respectively) bear some relation (X and Y) to the inferred claim (or *tercium comparationis*, TC following Spooren (1989) and Lagerwerf (1998)).

Notice that X and Y do not have to be identical or even opposite. This relaxed model of concession allows for examples in which speakers argue via different means (e.g., by presenting evidence in $P1$ and denying (via e.g., an anti-conditional, following RST terminology) in $P2$. So we view the two turns communicating the relations $X(P1,TC)$ and $Y(P2,TC)$ respectively, where the relations X and Y are inferred from $P1$ and $P2$, and we model cross-turn concession as $concession(X,Y,TC)$. This means that the form of a concessive dialogue is as follows, where we indicate what is being communicated and also whether it is necessarily explicitly or implicitly communicated in each turn:

(6.2)
A: $P1$ is explicit and $X(P1, TC)$ is implicit.
B: $accept/partial-accept(B, P1)$ or $agree/partial-agree(B, P1)$ is either explicit or implicit; "But $P2$" is explicit and both $Y(P2, TC)$ and $concession(X, Y, TC)$ are implicit

Since we take a subjective perspective towards interpretation in dialogue and we are focusing on modelling the speaker of the "but", we will view this concession as being communicated by B.

We will start by introducing cross-speaker concession in TOD and attempt to classify certain types of concessive phenomena in order to address what is actually being communicated. We will then distinguish different cases based primarily on Speech Act (SA), but we will argue, following Poesio and Traum (1998) and others, that the SA wraps around a central proposition, and we will attempt to distinguish cases based on properties of this central proposition as a means of generalising our treatment to address different SAs. Finally we will present a procedure distinguishing concession which enables what is communicated in the various situations to be modelled in the PTT (Poesio and Traum, 1998) model of dialogue. The goal of this work is to enable dialogue systems to distinguish situations in which different concessive phenomena are communicated and update their Information State (IS) representations of the dialogue accordingly, enabling better response generation.

6.1 Concession in TOD

Having discussed briefly how approaches designed with monologue in mind will not work across turns in dialogue, we will start here with the premise that traditional SA and planning information also fail to predict what is communicated by cross-speaker concession in TOD, since they ignore the "but" triggered inferences w.r.t. the TC. Therefore we start by focusing our analysis on what is communicated at the planning level in examples from the Maptask and TRAINS TOD corpora[4] and argue for criteria which a procedure that interprets concession and generates an interpretation based on

[4]MAPTASK dialogues can be browsed interactively at www.ltg.ed.ac.uk/~ amyi/maptask/demo.html and TRAINS dialogues can be found at www.cs.rochester/research/trains/. See Chapter 3 for more discussion on corpora.

the presence of certain salient (predominantly plan-related) features must have. Due to the sparseness of cross-speaker concession data involving "but", we model two examples from the TRAINS and Maptask TODs rather than model a statistically frequent type of cross-speaker concession.

We argue that semantic and SA representations do not provide enough information to distinguish what is really being communicated at a task level, and motivate our work by claiming that understanding both task-plan-related communication and semantic inferences is essential for full interpretation and generation of utterances in TOD. Our treatment of concession extends Lagerwerf's treatment of concession to dialogue, and connects concessive cases in TOD to the planning information necessary to interpret them.

6.1.1 Characteristics of Cross-speaker Concession

We start by determining what criteria are salient for distinguishing concession in a constructed example illustrating the sort of situation we initially consider, and then analyse examples from the Maptask and TRAINS TOD corpora. We then determine how concession might be generated in the IS framework and focus on how the utterances connect with planning operations in the task, employing plan-recognition models like those described in Litman and Allen (1987) and the Collagen project (Lesh *et al.*, 1999). Our approach follows the concessive analysis presented in Lagerwerf (1998), extended to TOD, where we will connect the semantic interpretation with salient planning information. Lagerwerf claims that concession requires a contextually available claim (or TC), for which both a positive and a negative argument are provided. Extending this approach to TOD, we argue that the positive and negative arguments for and against the TC are (1) not necessarily positive and negative arguments only, but simply different relations (i.e., X and Y in the scheme above), and (2) are entailed as defeasible expectations launched by the relevant assertions in the adjacent speaker turns in these cross-speaker "but" cases with respect to the TC. We will determine criteria for predicting concession in order to interpret the TC and stance of the speakers (w.r.t. the TC) in cases where interpreting concession is reasonable.

Let us consider how concession might be interpreted in actual cases by tracing

how we might analyse a simple example first. The concessive interpretation procedure proposed in section 6.3 describes four particular situations. In the first case, there is an opening question about which both speakers debate, as in the example below:

(6.3)

B1: Should we add the mushrooms to the sauce now?
TC = we should add mushrooms to the sauce now

A2: <That should be fine.> p
$X = agree/accept(p, TC)$

B3: (Ok) But <I thought we need to add the beans first.> p
$Y = introduce\_preceding\_action(p, TC)$

Here the TC is the proposition which is questioned in B1, and A argues for it by accepting the task action described in the proposition, while B argues against it by introducing an action which must precede the action contained in the TC. Notice that here X and Y are not opposites, but by referring to what they propose doing in the task-plan, we can see how they are possibly in conflict about what to do next. In fact, the temporal ordering of the planning operators in this example fits nicely within the concessive interpretation, since an opposing perspective is presented while possibly conceding the previous turn. Although B3's acceptance or rejection of the TC is not clear yet, she probably accepts adding mushrooms, or she would have rejected it initially and not proposed doing something else *first*, implying that the mushrooms can be added afterwards. It is also possible that B raises the question because of her own doubts about the validity of adding mushrooms at this point in the task, and wants to debate the issue with A in order to resolve what to do. Her doubt is emphasised by the objection she raises in B3. A would then need to respond to this objection by either denying B's concern that beans need to be added first, or by agreeing to B's implicit proposal to either add the beans first, or not add the mushrooms to the sauce.

We will see lots of examples in TOD in which differences in temporal ordering of planning operators in the speakers' respective task-plans lead to debates that can be interpreted concessively. Notice that without taking into account the speakers' task-plans, this example looks very polarised, with A supporting and B opposing the indirect suggestion of adding mushrooms at the current task-stage. This is because the TC is couched in a yes/no question, which is reasonably answered by agreeing or disagreeing

to the questioned TC.

Another case which the procedure addresses involves A proposing a plan or goal (as in the example below), and B objecting to this goal:

(6.4)
A1: <Let's add the mushrooms to the sauce.> p
TC = we should add the mushrooms to the sauce, $X = propose(p, TC)$.

B2: But <don't we need to add the beans first?> p
$Y = introduce\_preceding\_action(p, TC)$.

B2′: But <the sauce is already too thick.> p
$Y' = precondition\_failure(p, TC)$

Notice here that B2′ is an evaluative statement, and could equally well argue for the overall task-goal of making a good sauce. In fact, any parent or ancestor (i.e., higher level goal that the current goal is a part of) node in the planning hierarchy will provide a reasonable potential TC. The one given in the example above is the lowest level goal applicable, and therefore the most specific. Although we will only touch upon this briefly here, we advocate trying the lowest level goal as the TC before trying higher level ones, as this will result in much more specific interpretation. For example, the top level goal of making a good sauce will always be applicable in this TOD, while the lower level ones will not, as they are much more context dependent, simply due to the hierarchical nature we assume of the task-plan.

In Example 6.4 above, A1 introduces the TC by proposing it, which, although different from the previous example, also fits the concessive scheme we propose. Notice also that it will be necessary to interpret task-actions communicated w.r.t. the speakers' respective plans in order to determine that B2′ is providing a reason w.r.t. the TC against A1's proposal, i.e., by pointing out a negative effect which will be exacerbated by A1's proposal of adding something more to the sauce.

6.1.1.1 Analysing an Example from Maptask

Maptask dialogues[5] involve two participants with slightly different versions of a map with various landmarks on it, and the goal is for the route giver (who has a route drawn on her map) to direct the follower along this route from start to finish, so a large part of

[5] Discussed in more detail in Chapter 3.

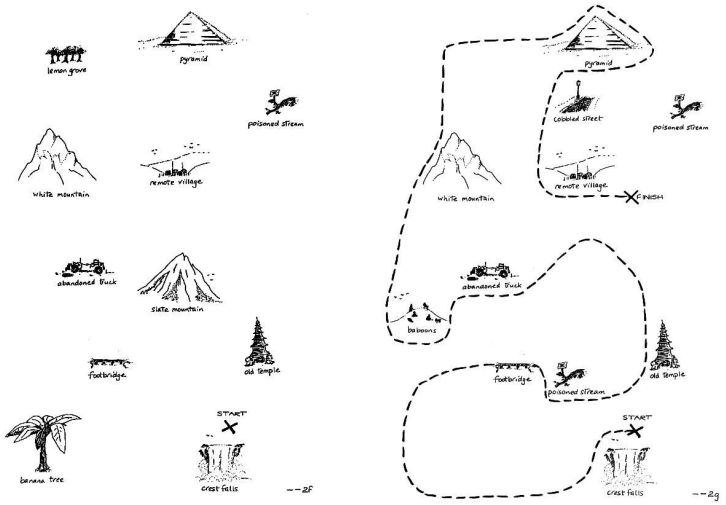

Figure 6.2: Follower map (left) and Giver (right) map for Example 6.5

the task involves establishing that the other participant has the same salient landmarks along the route in order to communicate clear directions. This is the sort of situation assessment we see going on in Maptask dialogue q3nc4 below (with disfluencies removed):

(6.5)

F1: right so i'm going underneath this mountain and along to the abandoned truck ... not underneath the mountain ... above the mountain?

G2: above the mountain yeah.

F3: so it's to the right of the mountain that i'm going?
TC = we should take the path to the right of the mountain

G4: <the abandoned truck's to the left of the old temple.> q
$X = clarify(q, TC)$

F5: yes, BUT <the mountain's between the abandoned truck and the old temple> p.
$Y = constraint(p, TC)$

The DAMSL annotation scheme (Allen and Core, 1996), which annotates utterances by coding their multi-level direct SAs, would predict that G4 has a Forward-Looking Function (FLF) of an assertion, and F5 has a Backward-Looking Function (BLF) of partial-acceptance (under the agreement category) and also has a FLF of

assertion. In terms of planning operator type, both G4 and F5 are static facts.

Neither semantic information nor direct SAs nor planning operator-type seem to shed much light on the issue of how concession might be interpreted here. Clearly we need to delve deeper into the actual problem the agents are trying to solve, namely their plans, in order to get at the underlying meaning of what is being conveyed by F5's "but". The follower (F) has the mountain on her map which the giver (G) does not have, and G only realises this discrepancy in maps after hearing F5.[6] This portion of the dialogue establishes the environment (and therefore path) that must be followed in this part of the map. I.e., the overall goal is to establish the route that must be traversed, which involves the subgoal of establishing salient landmarks around the given route that must be circumnavigated.

Concessive analysis

F5 can be seen as expressing concession, since F agrees with G4's assertion but asserts previously unmentioned information which is concessive because it indicates that despite G4's assertion being acceptable, G4's attempt to respond to her question (F3) has not answered it. Notice also that F3 indirectly proposes a route around the mountain, and that G4's indirect answer (i.e., assertion of a fact) can only be interpreted with respect to F3, since G must believe that the orientation of the truck with respect to the temple helps F to recognise the route she is trying to navigate. F5 accepts this information, but implies that it fails to answer F3 via a "but"-cued assertion. That is, F5 points out that there is a mountain between the truck and the temple. This assertion then needs to be checked against G's plan in order for him to determine where the contrast lies. In this case, the plan (which includes relative locations of objects on the map as constraints and preconditions of navigational actions) indicates that the path needs to be defined with respect to the mountain (i.e., F needs to know whether to go above or below the mountain).

The TC in this case can be framed as the proposal raised in F3 to go to the right of the mountain, where G4 has misinterpreted which mountain F refers to and indirectly corrects F by trying to point out the salient landmarks, while F5 accepts what G4 asserts (i.e., that the truck is to the left of the temple), and introduces new information

[6] G2 thinks F1 refers to the white mountain (a different one), which is further along on the route, and responds accordingly.

indicating that the mountain she asks about lies between the truck and temple. This is a very difficult example to analyse, and involves several steps of reasoning before it becomes clear how the propositions in G4 and F5 relate to the TC. Given this analysis, X would probably be correction while Y would introduce the new constraint (of where the mountain lies).

Also the concession expresses acceptance of G4's assertion but indicates that this assertion fails to answer the question, which is a Speech Act (SA) level relation rather than something at the planning level, but planning information is necessary to determine this. Determining where the contrast lies with this planning information also enables detection of discrepancies and deficiencies in their maps and facilitates alignment of their perceived environments. After hearing F5, G understands F's difficulty (they have aligned their maps), and communicates the route to F taking into account the mountain that is on F's map.

Notice that thinking of the two turns as supporting and denying the TC does not add much useful information here, and is not even really accurate, since the speakers are not disputing a claim but resolving a confusion; it is more useful instead to think of the two turns as expressing different perspectives on the environment, which is facilitated by X and Y in the interpretation scheme presented earlier. Interpretation will enable the route to be specified with respect to the mountain that lies between the two features (i.e., the truck and the temple) that G mentions.

6.1.1.2 Analysing the TRAINS example

In the TRAINS TODs[7], the agents' plans are considerably more complicated than in many other TODs (e.g., Maptask) and involve planning future actions rather than interleaving task and speech actions. Here the agents plan how to achieve a transportation task together where one agent knows additional constraints and guides the user. In the dialogue subsection below[8], one needs to understand where in the plan the speakers are in order to recognise that the user (U) is actually proposing a new subgoal with her assertion of a precondition of this subgoal which is achieved, and that the system (S)

[7] See www.cs.rochester.edu/research/trains/ and Chapter 3.

[8] From d.93-15.2, with disfluencies removed.

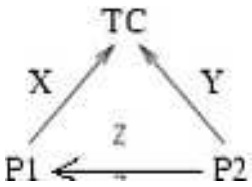

Figure 6.3: Concessive interpretation schema with related turns

agrees that this effect holds but rejects the validity of the indirectly proposed subplan (on the basis that the desired task goal, in this case getting the maximum amount of oranges to Bath by 7 a.m., would not be met).

(6.6) U1: well actually <I'll already have an engine in Bath after I unload the boxcars> p
right
TC = plan is OK; $X = precondition\_met(p, subgoal(TC))$

S2: right but <you wouldn't have it in time> q
$Y = precondition\_failure(q, subgoal(TC))$;

Here the TC would be that the plan is OK, with X giving evidence in support of the TC and Y accepting the validity of this information but rejecting it as evidence in support of the TC by denying that U1's assertion would meet the preconditions of the next stage in the plan. In particular, X indicates that a necessary precondition of a subgoal for the plan's success is met, and Y indicates that this precondition fails. This example is particularly difficult to analyse because it involves an implicit subgoal proposal in X. Notice that S2's turn involves both providing negative evidence towards the TC (i.e., that the plan is unlikely to succeed) and also rejecting the evidence given by U1. This adds an extra directional link between S2 and U1 such that S2 rejects the evidence given by U1. This means that in this case, we modify the scheme given in Figure 6.1 by adding a directional link (the relation Z) from S2 (P2) to U1 (P1), as illustrated in Figure 6.3.

Where semantic and SA analyses fail

Semantic representations of the utterances clearly cannot capture the inferences that must be made in order to determine that U1 is actually indirectly proposing a plan. Considering the SA level, the DAMSL annotation scheme would predict for U1 the FLF of information-request and assertion and the BLF of reject-part (cued by the "ac-

tually"). (Indirect SAs are not marked, though if they were, then the utterance would be annotated as an action-directive.) DAMSL would classify S2's utterance as involving the BLFs of both partial-acceptance and partial-rejection (for the "right" and the "but you wouldn't have it in time" parts of the utterance respectively), along with the FLF of assertion.

Simply mapping the assertional content of the turns to planning operators maps both turns to effects in this case, which does not give enough information to infer that U1 is proposing a new plan. Clearly we need to account for where in the joint plan these operators occur and how they are related, i.e., we need to account for prior planning actions when analysing cases like this, which we argue can be facilitated by accessing a task-plan history (TPH) which keeps a record of planning actions that have occurred so far in the dialogue, and which can be maintained in a dynamic semantic representation like the IS framework.

Where plans can help

Example 6.6 involves an implicit plan proposed by U1 to take the engine already in Bath back up to Corning with the boxcars to collect more oranges and return to Bath. This implicit proposal is recognised by S2, who then criticises the effect of this plan with respect to achieving the agents' main goal of getting to Bath by 7 a.m. with the maximum amount of oranges by indicating that they cannot get the oranges back to Bath by 7. In our approach, we assume that a task action (TA) interpreter communicates with the planner, and that this TA interpreter can recognise the implicit proposal in U1's assertion and the criticism of this proposal's effect in S2's turn and keeps the IS updated with what is happening in terms of the agents' salient planning actions which are stored in the task-plan history (TPH) field (both the TA interpreter and TPH will be discussed more in the next subsection). The TA interpreter maps utterances onto planning operators and communicates with the planner (e.g. by forward or backward chaining on the planning operators mentioned) to determine potential conflicts etc. This is feasible in restricted TOD domains and has been implemented in other systems (e.g., the BEE tutorial dialogue project[9] and the Collagen project (Lesh *et al.*, 1999). Notably Blaylock and Allan (2005) present a model of collaborative problem solving

[9] See www.cogsci.ed.ac.uk/~ jmoore/tutoring/index.html.

which is built upon problem-solving (PS) objects which are typed feature structures that themselves model planning operators. Their abstract PS objects are things like objectives, recipes, constraints, evaluation, etc. and enable them to explicitly model both agents' decision-making and the process they follow (e.g., evaluating multiple recipes) to make decisions, and they argue for the need to account for both. Furthermore their collaborative PS states represent the evolving status of various PS objects, e.g., whether they are adopted, accomplished, etc., which effectively represents the sort of information produced by the TA interpreter and found in the TPH. It should be noted that this approach is compatible to ours. However this approach integrates planning within the dialogue model, while we opt for the alternative approach of outsourcing our planning operations via the TA interpreter, as planning is beyond the scope of this thesis.

6.1.2 Incorporating Task-Plan History in the IS

We saw in Example 6.6 above that in order to determine the TC we needed to infer U1's proposed plan and record it in the IS, requiring a dynamic representation of plan history for the dialogue so far. In the last subsection we introduced the idea of a task-plan history (TPH) field in the IS which keeps a record of planning actions that have occurred so far in the dialogue, and which are dynamically updated whenever the IS is updated. The idea was for a task-action (TA) interpreter which interfaces with the planner to take conversational actions (i.e., CAs, as they are updated in the IS) and determine via the planner and a model of the speaker's task-plan and current stage in the plan which task-actions are being communicated. The TA interpreter determines (for example) whether two operators are alternatives, whether they conflict, what the consequences of a given action communicated are, what its effects and preconditions are, etc. Its role is to extract the planning information communicated in TOD turns relevant to the speaker's evolving plan in order to get at what is really communicated in TOD turns. The TA interpreter then updates the IS with a record of the task-plan actions corresponding to conversational actions which are communicated at each turn. This information about task-actions communicated gets stored in the task-plan history (TPH), the analogue to the dialogue history (DH) which stores conversational actions communicated. For example, if an utterance "I think we should take the boat to Port

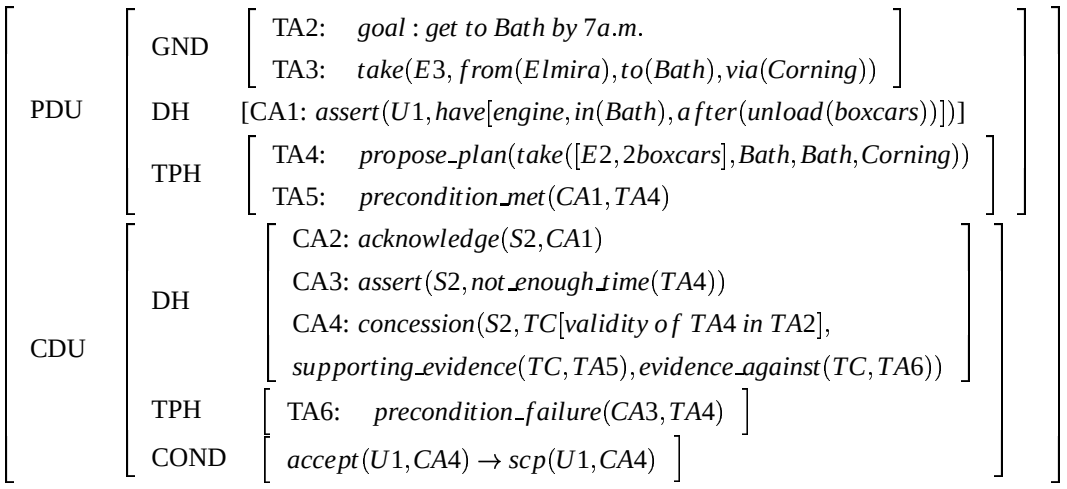

Figure 6.4: The IS after S2's turn in the concessive interpretation of Example 6.6

Hope" is uttered, the CA interpreter (discussed more in Chapter 7) translates this to the CA: $suggest(take(boat, to(Pt\ Hope)))$. The TA interpreter has a pointer to the current stage in the task-plan, and resolves referents (e.g., the boat) to trace what has happened with these referents so far and determine their context. The TA interpreter then comes up with a list of hypotheses, e.g., things like $goal : take\ boat\ to(Pt\ Hope)$. These then get updated in the TPH as TAs. We do not address the workings of the TA interpreter further, as it is beyond the scope of this thesis, but we point to Blaylock and Allan (2005) and projects like BEE and Collagen to indicate that this is feasible. Figure 6.1.2 shows an example of the TPH field in the IS.

Consider the IS shown in Figure 6.3 which contains salient task-plan information that helps us interpret the TRAINS example (6.6). The IS is broadly grouped into Previous and Current Dialogue Units (PDU and CDU respectively), where U1's turn is contained in PDU and S2's turn is in CDU. Previous conversational acts (CAs)[10] are stored in the Dialogue History (DH), and we store previous task actions in the TPH, both of which show acts in chronological order. The Common Ground (GND) holds information which has been accepted by both speakers, and the Conditions field (COND) holds $if-then$ rules which in this case indicate how speakers are socially committed (Kreutel and Matheson, 2000) to the concession relation if they accept it. We omit

[10] See Chapter 9 for more information on the PTT (Poesio and Traum, 1998) model of dialogue which we use.

irrelevant fields and acts here for brevity, and just show what the relevant part of the IS (following the IS structure given in Matheson *et al.* (2000)) would look like for U3 (after hearing S2 utter CA4). Notice that TPH contains Task Actions (TA), mirroring Conversational Acts (CAs) in the Dialogue History (DH). The recognition of CAs as signalling specific TAs (e.g., TA4 and TA5) would occur within a TA interpretation process which is called by the dialogue manager, and would interface with the planner, often involving plan-recognition to determine what is being communicated, (e.g., Litman and Allen (1987) and Lesh *et al.* (1999)). In these cases the interpreter would need to call the planner with the specific planning operator communicated and have the planner forward-chain from this operator to see if the TC is achievable given this input. These processes would be called by update rules in the dialogue manager that call the TA interpreter in order to determine what intentions are being communicated in the utterances.

For Example 6.6, before update CA3 would contain $assert[S2, not\_enough\_time(do(Z))]$; in the IS above, the argument Z of the anaphoric action communicated by S2 in CA3 has already been resolved to TA4 in the update process by recognising that U1 is making a proposal. U1's indirect plan proposal can be inferred via the interpreter, planner and the TPH; i.e., TA4 is inferred by the planner upon receiving CA3, thereby enabling recognition (in the planner) that an alternative to TA3 is being proposed, and resulting in TA5. This resolution should occur before the update module determines whether or not to generate concession, and the procedure proposed in the next section takes as input an IS that has been updated with this information, which simply requires the dialogue manager to call the procedure after applying the update rules which interpret the input and updates the IS with the TAs recognised by the planner. So to resolve anaphoric actions like $do(Z)$ in CA3, we need to incorporate into the update module a rule of the form:[11]

- If PDU.TOGND.TPH contains a TA that matches $propose\_alternative(TA_i, TA_j)$ and if PDU.TOGND.DH contains a CA_n of the form $assert(Y)$ where Y contains $do(X)$, then replace $do(X)$ with TA_i in Y.

[11]This rule is an example of many similar rules for other situations involving anaphoric reference to actions that need to be resolved in order to interpret the dialogue.

The Benefits of Modelling Cross-speaker Concession

If U1 accepts the concession communicated by S2, e.g., by saying "Yeah, I guess you're right, going back up to Corning for more oranges wouldn't work", then she is socially committed (Kreutel and Matheson, 2000) to the concession relation being communicated with the given TC and S2's expectation that her proposed plan would not work. In fact, she needs to recognise that S2 is responding to her implicit plan in order to know that S2 recognised it. On the other hand, if she does not accept S2's concession, she can indicate this by saying "But it only takes an hour to travel between Bath and Corning, so I should get it there in time", which requires that she understands that S2 is signalling a rejection of the implicit plan she proposed, and counters S2's argument that her proposed plan would not work. Our approach addresses cases like the Maptask and TRAINS examples via update rules in the dialogue manager which call the interpreter. The interpreter then passes the TAs it interprets back to the update rules which update the IS accordingly with the given TAs. An update rule in the dialogue manager triggered by "but" then calls the concessive interpretation procedure, which tests whether the given situation can be interpreted as concession. So both recognising these implicit plans and also the concession relation S2 communicated is essential for her to generate an appropriate response in cases like these.

6.2 Classifying Concession

In this section we will consider concession in both NTOD and TOD from the perspective of Speech Acts (SAs); that is, we will examine concessive responses to commands, questions and assertions in both TOD and NTOD to gain intuitions about the range of types of concession we can encounter. In the last section we saw that concession in TOD signals differences in goals, actions and plans; in contrast, concession in NTOD signals a difference in perspective, opinion or belief. Bearing differences in domain in mind, we will now turn our attention to the more general trends encountered in concessive responses to SAs.

Although we will use SAs to distinguish between the various cases of concession as a starting point, we will follow Austin (1962) and others in perceiving SAs as il-

locutionary wrappers around a central proposition. This crucially means that despite approaching different SAs which trigger concession separately, it is our aim that what will emerge is a central treatment for propositions with differences in modelling arising to account for SA differences. Significantly this perspective is in accordance with the PTT model of dialogue (Poesio and Traum, 1998), facilitating our modelling attempt. More generally this perspective enables a more modular treatment of SAs and furthermore separates characteristics of the central proposition from those of the SA, allowing a clearer distinction of which aspects of utterances contribute to concessive interpretation. In the discussion that follows, we will consider SA as the primary distinguishing aspect, and after considering the contributions of different SAs, we will determine which aspects, illocutionary or propositional, contribute to concessive interpretation.

We will draw upon aspects in common in the different cases examined to generalise about how concession relates material across turns. As shown in the concessive scheme in Figure 6.1, we will find relations X and Y between the TC and the propositions in the turns respectively, and occasionally also a relation Z between the two propositions themselves. We will see that in all these situations, the concessive turn accepts or minimally acknowledges the prior turn while presenting an alternative argument w.r.t. the TC. Due to the sparcity of data containing cross-speaker "but" expressing concession and the lack of a systematic method of extracting these cases from the corpora, the data in this chapter, as in the majority of this thesis, is predominantly made up. While this is not ideal, we hope that the plausibility of these examples speaks for their relevance and frequency in spoken dialogue.

6.2.1 Responding to Commands

Commands which are responded to concessively will be considered TOD, since they involve agents performing actions and notions of task-plans will help resolve these cases. In the example below, the TC is the state of washing the dishes (with no illocutionary force) and X commands the accomplishment of the action of washing the dishes, which results in the state in which the dishes are washed. The commandee then addresses the achievement of the TC, hence the $achieve(TC)$ in all of the commandee's

possible concessive responses (B1-B6 below):

(6.7)

A: <Wash the dishes.> p
TC=you should wash the dishes; $X = command(p, achieve(TC))$

B: But <we haven't had dessert yet> p
$Y = constraint(p, achieve(TC))$

B1: But <we did them earlier> p.
$Y = invalid\_action(p, achieve(TC))$

B2: But <we don't have enough water to wash them.> p
$Y = precondition\_failure(p, achieve(TC))$

B3: But <we should clean the house instead.> p
$Y = alternative\_action(p, achieve(TC))$

B4: <But what else do we need to do?> p
$Y = question\_alternatives(p, achieve(TC))$

B5: <But where will we find the detergent?> p
$Y = question\_precondition(p, achieve(TC))$

B6: OK, <But first give me some detergent!> p
$Y = request\_precondition(p, achieve(TC))$

In response B, the speaker presents an objection to the command in the form of a constraint for the action A commands that one has finished eating before washing the dishes and does not disagree with the goal of washing the dishes. B1 accepts the validity of the goal behind A's command, but argues that the command cannot be performed because the desired state has already been reached. B2 presents a precondition failure for accomplishing the desired action and also communicates only implicit acceptance of A's goal of reaching a state in which the dishes are washed. B3 presents an alternative action and also communicates just implicit acceptance of A's goal. B4 only expresses understanding of A's goal, not necessarily accepting it, and asks about alternative goals that can be pursued. B5 asks about how to achieve a precondition to the commanded action. Notice here that the *what* question is not easily reduced into a central proposition being queried, while the *where* question is much more easily reduced into a central proposition, and it is much simpler to strip away the SA wrapper and frame it as (for example) an assertion, like B5′: But we can't find the detergent. B6

commands the accomplishment of a precondition (being given the detergent) before he can wash the dishes.

Notice that for all of B's responses above, B never explicitly refuses to perform the commanded action. It seems reasonable to argue that in "but" cases one cannot respond with a "no, but" or equivalent negation followed by the "but". This should hold because, as in monologue, "but" is logically equivalent to "and", which cannot deny one of the premises conjoined to the other. Consider B7: "I can't wash the dishes but I can dry them". This is fine as a rejection of the command, with the intra-turn "but" softening the rejection by offering another form of help. "I can't wash the dishes!" and "No!" also work as flat rejections. Even B8: "no, but I can dry them" is reasonable. So how do we account for this oddity? If "but" is logically equivalent to "and", then it cannot reject something. We can argue that on a meta-level, "but" in this case is indicating something additional, rather than rejecting the previous turn's command. Consider "no, and I'll tell you why: I have a cut on my finger" as a response. This also works fine despite containing "and" itself, since "and" is linking the performative on a meta-level to the rejection as an explanation.

Situations in which speakers can phrase the intended meaning behind their proposition equally effectively as a question or an assertion will be interpreted as an assertion if the question requires indirect SA interpretation. E.g., in B3 in Example 6.7, B could just as easily ask, "But shouldn't we clean the house first?". In terms of concessive analysis, there is no practical difference between this question and its almost equivalent assertion "But we should clean the house first". Despite not making a significant difference in our analysis in this case, SA can affect the response that the hearer might make. In this case, as in many others, a question asks for A's opinion, while if B uttered the equivalent assertion, she does not allow A to negotiate with her nearly as much as if she frames her opinion as a question. This is possibly also why the question form seems more polite than the assertion; it leaves open the possibility of discussion, since B presents her opinion (which is an alternative to A's) and then waits for A to respond.[12] Notice that wh-questions are not as easily mapped to an equivalent assertion; hence the questions in B4 and B5 can only be interpreted as questions and have

[12] We do not account for other "softeners" like those that explicitly indicate opinion or perspective, hesitation, etc, e.g., "I think...".

no equivalent propositions which are assertions or commands.

6.2.2 Responding to Questions

Questions in NTOD and TOD are often followed by either concessive or corrective dialogue, the latter of which will be discussed in Chapter 7. However, we also see, in both TOD and NTOD, concessive dialogues following questions which do not involve correction, e.g., proposal of alternative options in TOD. These often follow responses to questions rather than the question themselves, but we will address them in this subsection rather than in the subsection addressing responses to assertions, as the opening question focuses the debate and provides the TC in these cases. In fact, these concessive responses to answers/responses to questions require the opening question to be interpreted, and can often be seen as providing an alternative answer/response which contrasts with the former response. We will focus on these responses to answers before turning to concessive responses to questions themselves.

(6.8)

A1: What should we do next?
TC = the next good action to be pursued is X

B2: Let's <go to the abandoned cottage.> p
$X = propose(p, instantiate(TC))$

A3: Ok, but <we could save time by going directly to the mirror of tears.> p
$Y = propose(p, instantiate(TC)); Z = propose\_alternative(p, X)$

A3′: Ok, but <first we should stop at the stone well to pick up the golden key.> p
$Y = propose(p, instantiate(TC)); Z = introduce\_precondition(p, achieve(X))$

In A3′ and A3′′ the speakers introduce preconditions which must be achieved before the TC can be achieved, and they (A3′ "wins" by presenting arguments which respond to B2's proposal but introduce new information. A3 acknowledges B2's proposal but proposes and argues for an alternative course of action to achieve the TC. In this case, the scheme described in Figure 6.1 can be modified to hold an additional relation between X and Z, since Z objects to X by presenting an alternative. In A3′, speaker A accepts B2's proposal, but qualifies it by introducing an action which precedes it. In this case her proposal wins by virtue of apparently not conflicting with

B2's proposal. The proposal of goals and actions makes this TOD rather than NTOD.

Notice that the question raised by A1 is not one that can be verified as true or false at the time that she utters it; instead, A1 raises the question in order to determine what the future course of action should be, which results in both agents proposing and debating various plans of action. Another question which should be asked is whether it is only these opening questions which can be responded to concessively; i.e., does concessive dialogue hinge upon an explicitly raised question which encourages discussion of alternatives? Consider the following dialogue in which the opening question asks for a specific answer:

(6.9)

A1: What time is it?
TC = the current time is X

B2: <It's 4 o'clock.> p
$X = answer(p, question(TC))$

A3: But <that can't be true!> q <It was just 1 o'clock a few minutes ago!> p
$Y = answer(p, question(TC))$; $Z = deny(q, X)$

A3′: But <when was our train supposed to leave?> p
$Y = question(p)$; $Y' = question(p, given(X))$

A3″: But <I thought it was 2 o'clock.> p
$Y = answer(p, question(TC))$; $Z = propose\_alternative(p, X)$

This example involves a wh-question which is answered by B2. B2 is then contested by A3 and A3″. A3′ raises a new question that is either unrelated to the TC or assumes that X (B2's answer) holds. A3′ cannot be interpreted as concessive, because, as we can see, there is no relationship between what is uttered and the TC. Therefore it does not fall into the scheme of cross-turn relations we deem concessive, which must adhere to Figure 6.1 or 6.3.

Following Larsson (2002), the question A1 raises a question under discussion (or QUD, Ginzburg (1996)) which is then answered by B2. The QUD is seen as part of the speakers common ground (Clark, 1992), and the responses to the answer in B2 (i.e., A3, A′ and A3″) all indicate to different degrees the acceptance of the answer B2. The concessive responses A3, A′ and A3″ indicate grounding problems at different levels, where all A's responses accept that B2 is attempting to answer A1, and A3 denies the

truth of this answerhood, with A3″ also expressing doubt about the truth of B2 but not denying B2 like A3, while A3′ accepts B2 and asks a follow-on question which explains the relevance of her initial question A1.

Question-type plays a not-insignificant role in determining what must reasonably follow the question turn, as we will see in the chapter on correction. In fact, we will see that questions followed by a pair of assertions can often be interpreted both concessively and as correction. *What*, *where* and *when* questions all expect a single value that is being asked about to be answered, and as such, in cases involving the question followed by two debating turns, it would be natural for the speakers to debate over this value. However they can also criticise the question's relevance or propose an alternative suggestion. In the TOD example below, the speakers debate over the next ingredient to add:

(6.10)
B1: What should we add to the sauce next?
TC = we should add X to the sauce next

A2: <Let's add mushrooms.> $p\ X = suggest(p, TC)$

B3: OK, but <shouldn't we add beans first?> $q\ \ accept(B, A2)$, $alternative\_proposal(q, TC)$

How and *why* questions are a bit different; the former expect alternative methods and the latter alternative reasons, as in the following two examples:

(6.11)
B1: How should we prepare the vegetables?
TC = we should prepare the vegetables by doing X

A2: <We should mash them.> $p\ propose(p, TC)$

B3: Ok, but <first we should wash them.> $q\ propose(q, TC)$

(6.12)
B1: Why do dogs eat meat?
TC = dogs eat meat

A2: <Because they evolved that way.> $reason(p, TC)$

B3: (Maybe,) but <they also just prefer it probably.> $q\ propose(q, TC)$

Notice that in both these cases the turns need to be recognised as proposing alternatives in order to interpret that they are concessive. The responses all show understanding that B1 is asking a question, and so to be cooperative (Grice, 1975) they propose

answers which respond to a question that asks about the TC. For Example 6.9, A3″ proposes an alternative to *X* (i.e., B2) as an answer to this question, where we again get the triangular variant of the concessive scheme illustrated in Figure 6.3 with the addition of the *propose_alternative* relation between A3″ and B2. Notice that all the A3 responses also take into account B2's answer, hence for A3′ we also get the alternative relation Y′, where the same question that is asked in *Y* is asked, given the additional information in *X* (B2), which could be responsible for triggering A3′ to ask when their train was supposed to leave. Y′ can be viewed as an alternative interpretation to *Y* for A′.

A3 and A3″ are ambiguous between correction and concession because both involve denying the previous turn's answerhood (implicitly in the case of A3″) while proposing a different answer. They can also be interpreted as concessive since we assume that they demonstrate understanding of *X* (in B2's turn).

Wh-questions generally (with the exception of "what should we do next" types of questions) ask about a particular attribute, in this case time, which is then answered with a value, which can then only be directly contested. That is, they do not open up possible alternatives the way the "what should we do next" question above did.

Planning issues do not occur in NTOD, and in concessive situations responding to questions, the questions must inquire about a situation that has a truth-value or known answer in the world at the time of uttering it. Alternatively, in NTOD speakers can ask about the other's opinion on some subject, which, for example, the askee may not know about; e.g., in the example below, if B2 does not know about X or their albums, he can say so. What is different about NTOD is that it will not pose questions requiring the agents to make choices for future actions. In NTOD questions can be raised and followed by a concessive dialogue that involves debate over opinions, perspectives or beliefs, e.g.:

(6.13)

A1: What did you think of the last X album?
TC = your opinion of last X album is Z

B2: <Oh I thought it was their most technically brilliant.> p
$X = answer(p, question(TC))$

A3: But <it sounded a lot like the previous album to me.> p
$Y = answer(p, question(TC))$; $Z = criticise(p, X)$

A3′: But <I think their first album was actually the most imaginative.> p
$Y = compare(p, TC)$;$Z = alternative(p, X)$; $Z' = criticise(p, X)$

A3″: But <when did they release it?> p
$Y = question(p)$; $Y' = question(p, given(X))$

Here denial of B2's opinion does not make much sense; instead A3 and A3′ introduce criticism of B2's opinion on the last X album by criticising the album under discussion, albeit indirectly in A3′, and presenting alternative viewpoints w.r.t. B2's opinion. A3′ discusses the band's first album, comparing it to the last one, and indicating that this album is better than their last one. Because this response also does not directly address the last album but compares a different album to the one referred to in the original TC, we model this as comparison. A3″ shows how wh-questions do not seem to be able to present an alternative stance here (unless there is a particular significance to the release date and both speakers know this and furthermore, know that the other knows this (Clark, 1992)). A3″, like A3′ in Example 6.9, can be interpreted in two ways, one of which is an alternative interpretation which takes into account B2's turn.

Barring the incorporation of task-plans to resolve TOD concessive responses to questions, one way of determining how the concessive response relates to both the question and the preceding answer might involve categorising questions conceptually, as done in question answering approaches, e.g., Lehnert's QUALM project, (Lehnert, 1977). This project involved 13 conceptual categories which enquire about: causal antecedent, goal orientation, enablement, causal consequent, verification, disjunctive alternatives, instrumental/procedural category, concept completion, expectational aspects, judgmental aspects, quantification, feature specification, and requests. Given a mechanism for distinguishing these categories, e.g., by parsing questions into these

categories, knowing the question category/type will indicate what sort of answer is expected, e.g., a question enquiring about a causal consequent expects an answer which provides this consequent. So knowing the question category/type will help determine if an assertion can be classified as an answer to the question. Then in cases in which an answer of the appropriate category type is provided, the concessive response can, assuming the speaker knows that the answer is incorrect or prefers an alternative answer, indicate this concessively. If the answer does not match the question category, this can also be indicated concessively.

One last situation which should be raised here is the situation in which a question is responded to with a concessive "but" assertion. This is the situation involving concessive response to a question itself, which we have not yet addressed. In the example below, we have a question which asks for the answerer's opinion (which would qualify as a "judgmental aspects" QUALM category) and expects that an opinion will be provided.

(6.14) A: <Didn't you like Ghost World?> p
TC = your opinion of Ghost World is X; $X = implicit\_answer(p, question(TC))$[13]

B: Yes, but <I wish it had followed the original graphic novel a bit more closely.> p
$Y = answer(p, question(TC))$; $Z = criticise(p, X)$

This example is classic concession, with B agreeing to A's implied opinion but raising a criticism about the TC as well. We interpret questions about speaker's opinions on a TC as before in Example 6.13 A3′, and we also interpret criticism in B's answer here.

Since in NTOD we cannot take advantage of a task-planner which can resolve how the planning operators communicated are related, we may not (barring theorem-provers and appropriate models for the speakers' private beliefs) be able to determine how B relates to A. However, in both NTOD and TOD, we can assume that there is some objection to A being raised in B's turn. Consider the following example:

(6.15)

A: <What should I add next?> p
TC = the next ingredient to add is X; $X = propose(p, question(TC))$

B: But <the sauce already has too many ingredients in it!> p
$Y = reject(p, action(TC))$

B′: But <we need to finish the sauce base first.> p
$Y = introduce\_precondition(p, action(TC))$

B″: But <the sauce is done already!> p
$Y = reject(p, action(TC))$

Here B communicates a problematic effect which will only be exacerbated if the agents perform A's proposal; i.e., B will produce an undesirable result/effect. The action of accomplishing the TC is what is being queried and responded to here. B′ introduces a precondition which satisfaction-precedes A's indirect proposal of adding something to the sauce. B″ informs A that her proposal is invalid, since the stage at which her proposal would be valid is past. This information can be determined via the task-planner,which should report (when called with the turns) what task actions are being communicated and how they are related in the speakers' task-plans. In these cases we want to capture what exactly is being objected to in B's turn.

6.2.3 Responding to Assertions

Assertions that are responded to concessively in TOD must often be interpreted in terms of planning operators communicated in order to determine what exactly is being debated, as we saw in the last section (e.g., Example 6.15). In NTOD however, conversations are not about achieving a task and communicating task-related information, but often simply involve the sharing of opinions, and can be interpreted without recourse to task-planning information communicated. Consider the following two examples:

(6.16) A: <Marilyn Monroe was the diva icon of the '50s.> p
TC = Marilyn Monroe was diva icon of the '50's; $X = support(p, assert(TC))$

B: (Yes,) but <she'd never be such an enduring icon without Warhol's use of her as a subject in his art.> p
$Y = criticise(p, reason(TC))$

B′: (Yes,) but <do you really think anyone would remember her without Warhol's paintings of her?> p
$Y = criticise(p, reason(TC))$

B″: (Ok,) but <then why did she disappear from the cover of Time magazine?> p
$Y = criticise(p, question(TC))$

B‴: (Sure,) but <where was she when Doris Day made her debut?> p
$Y = criticise(p, question(TC))$

Here we assume that A supports the TC by virtue of asserting it (i.e., it is this assertion of the idea of Monroe as icon that is supported). In B and B′, B agrees with A's assertion and introduces a qualifying factor of Monroe's fame; i.e., A asserts the TC, and B criticises the TC via the assertion that Monroe's fame $depends-upon$ Warhol's paintings of her. In this case, the dependence is causal, and we can represent this via the *reason* relationship, so that p is the *reason* for the TC to hold here. This is concession, since B's qualifying information limits the extent of A's claim. I.e., B agrees to A's claim, but limits its extent. B′ asks a question but communicates the same criticism. B″ presents a possible counterargument by introducing evidence indicating her lack of popularity at some later point in time. Since B″ asks why this was the case, we interpret this as p querying a criticism of the TC. B‴ asks about an alternative diva again as a form of criticising the TC. TOD can also involve similar examples. Consider the following example:

(6.17) A: <The sauce is ready.> p
TC = dinner is ready; $X = precondition\_met(p, achieve(TC))$

B: But <we don't have pasta.> p
$Y = precondition\_failure(p, achieve(TC))$

B′: But <no one wants dinner.> p
$Y = invalid\_goal(p, TC)$

Here we assume the TC of making dinner, i.e., the task goal. A's turn indicates that a necessary precondition for achieving the TC has been met, while B indicates that another necessary precondition cannot be met. B′ indicates that the TC is an invalid goal. Compare this to the following NTOD example:

(6.18) A: <Dogs are more faithful.> p
TC = dogs are better pets; $X = supporting\_evidence(p, assert(TC))$

B: (Yes) but <cats are more interesting.> p
$Y = evidence\_against(p, assert(TC))$

Here we assume the TC that considers whether or not dogs make better pets. Then we have X providing supporting evidence and Y providing evidence against this TC. This example clearly follows Lagerwerf's model for concession in monologue, where A and B provide alternate answers for the TC.

Arguments for and against the TC must be interpreted as indicative of speaker stance. Interestingly enough, while this example bears a strong resemblance to Semantic Opposition (recall examples like "Mary likes skiing but JoAnn likes chess" from monologue), it can also be analysed quite neatly as concession, given a relevant TC, as illustrated above. What does this say about the distinction between S.O. and concession? In monologue the distinction is much clearer, where there is no particular argument being made in S.O., and the two clauses are simply compared or held against each other. In this case, we saw that context enables the concessive reading, given an appropriate TC. However certainly this resemblance to S.O. is just coincidence in this example, since we have seen plenty of cross-speaker concessive cases which do not involve any visible opposition or comparison between turns. Presumably in some S.O. cases, given prior context in the form of a TC, a concessive analysis will also be possible, providing additional interpretation.

Notice also the different types of assertion-based concession dialogues presented; we need to distinguish between the cases supporting and refuting a TC like the previous one and cases which do not propose alternatives but instead introduce crucial qualifying information regarding the previous assertion. In the next section, we will address how these different cases are modelled in a procedure that updates the IS with what is being communicated in these concessive dialogues.

Summary:

In some cases (e.g., Example 6.7), the SA wrapper makes little difference and it is the content of the central proposition itself which is most involved in the concessive relation. This is in keeping with Austin's (Austin, 1962) perspective of SAs as illocutionary wrappers around a central proposition, and it is also in accordance with the PTT model of dialogue. We also saw that, barring differences arising from planning operations and their inter-relationships, TOD and NTOD do not behave so differently and domain did not make much difference in these cases. So what commonalities can we draw from this analysis? Well certainly none of these cases involved outright rejection. In some cases they might have put forward a favoured alternative, but it seems that one common factor is that all these cases involve at least acknowledgement if not acceptance of the prior turn, while providing different arguments X and Y w.r.t. the TC. While the pair of X and Y relations varied widely, the underlying features of acceptance or acknowledgement and the pair of differing perspectives w.r.t. the TC were in common for all the examples seen.

6.3 Modelling Concession

We propose an update procedure to model the cases of cross-speaker concession discussed thus far. The update procedure proposed would be called by an update rule in the dialogue manager triggered by "but". If the required circumstances (outlined in the procedure) apply in the given case, concession is predicted along with relevant inferences about the speaker's beliefs. We will model concession following the utterance of the concessive turn rather than address the generation of concessive utterances here, since to do the latter, we would need to take into account the speaker's other communicative goals, beliefs, etc. Our approach will enable the concessive relation to be accounted for in the IS in order to facilitate appropriate response generation to the concessive turn.

We will focus on the particular X, Y concessive pairs (and Z where applicable) to distinguish concessive situations. This distinction scheme bears in mind that SAs are essentially wrappers around a central proposition, since we do not make the distinction

based upon SA pairings of the concessive turns.

6.3.1 Concessive Operator Pairs

In this subsection we will summarise the concessive pairs for X and Y seen thus far in this chapter. Since we saw that these pairs of X and Y communicated cross-turn concession, we will interpret concession in other cases in which these pairings are seen, provided that the concessive turn acknowledges or accepts the previous turn. In the situations described below, Y_2 etc. simply indicate alternative Ys for a given X. X' and Y' indicate alternative X and Y possibilities that work in these situations.

- – $X = support(p1, assert(TC))$
  – $Y_1 = criticise(p2, reason(TC))$
  – $Y_2 = criticise(p2, question(TC))$

  Recall that this simply means that X supports the assertion of the TC while Y criticises either the reason behind the TC or questions the proposition expressed by the TC.

- – $X = agree/accept(p1, TC)$; $X' = propose(p1, TC)$
  – $Y_1 = introduce\_preceding\_action(p2, TC)$
  – $Y_2 = introduce\_precondition(p2, achieve(X))$

- – $X = precondition\_met(p1, achieve(TC))$
  – $Y = precondition\_failure(p2, achieve(TC))$

- – $X = supporting\_evidence(p1, assert(TC))$
  – $Y = evidence\_against(p2, assert(TC))$

  Both this case and the preceding one are quite similar and obviously concessive since they present the positive and negative aspects of the given TC. The former is only applicable in TOD and the latter in NTOD.

- – $X = propose(p1, question(TC))$
  – $Y_1 = reject(p2, action(TC))$
  – $Y_2 = introduce\_precondition(p2, action(TC))$

- $Y_3 = introduce\_preceding\_action(p2, TC)$
- $Y_4 = precondition\_failure(p2, TC)$

Although this pairing is not as obviously concessive, presumably what Y_1 rejects is what X proposes. Y_2 and Y_3 implicitly accept X's proposal but introduce preconditions and preceding actions. Y_4 indicates precondition failure with X's proposal.

- $X = implicit\_answer(p1, question(TC))$
- $Y_1 = answer(p2, question(TC)); Z = criticise(p2, X)$
- $Y_2 = compare(p2, TC); Z = alternative(p2, X); Z' = criticise(p2, X)$
- $Y_3 = question(p2); Y_3{}'' = question(p2, given(X))$

These cases involve the relation Z between X and Y as well, where Z criticises X or presents an alternative. Y_3 could overgenerate, since there is no obvious connection between what is questioned and X, so we will not model this case in the procedure.

- $X = clarify(p1, TC)$
- $Y = constraint(p2, TC)$

This case is not obviously concessive unless one indirectly interprets the clarification as a means of supporting the TC, but we will include it in the procedure since it occurred concessively in an example and concession itself is defeasible, so if it is incorrectly predicted, it can be corrected by simply denying that it holds.

- $X = command(p1, achieve(TC))$
- $Y_1 = constraint(p2, achieve(TC))$
- $Y_2 = invalid\_action(p2, achieve(TC))$
- $Y_3 = precondition\_failure(p2, achieve(TC))$
- $Y_4 = alternative\_action(p2, achieve(TC))$
- $Y_5 = question\_alternatives(p2, achieve(TC))$
- $Y_6 = question\_precondition(p2, achieve(TC))$

There are undoubtedly many more concessive pairs, and these are just the pairs we have found so far.

6.3.2 Procedure for Distinguishing Concession

Given the number of disparate cases in this non-exhaustive listing, we eschew modelling pairs of X and Y themselves as concessive. Instead we open up the interpretation to include all possible pairings of X and Y, and remove the onus of distinguishing concession from the pairings themselves. This means that we distinguish concession provided (1) the "but" turn accepts or acknowledges the previous turn and (2) introduces a different perspective (or relation X) w.r.t. the TC.

So the initial distinctions a procedure modelling these cases should make assumes that all pairings of X and Y predicates can be concessive and bases its decision on finding an X and Y in PDU and CDU respectively and also finding acknowledgement or acceptance in CDU. Assume a dialogue with B uttering "but" (i.e., turn i) and finding $P2$ (the "but" cued turn's proposition) and some relation Y w.r.t. some TC in CDU.TOGND.DH[14] and A's relation X w.r.t. the same TC and utterance $P1$ in PDU.TOGND.DH. Note that some variables (capitalised) in the procedure are uninstantiated; while this might allow overgeneration, we argue that this is preferable to not interpreting concession when it holds. That is, we will assume that overgenerating is better than missing too many cases, since what is overgenerated, i.e., concession, is defeasible, and the next speaker can correct it if they did not mean to communicate concession.

1. If current turn (B_i in CDU.DH) contains a turn-initial "but" and

   (a) If CDU.DH also contains [either an $agree(B_i, CA_p)$, $accept(B_i, CA_p)$ or $acknowledge(B_i, CA_p)$ CA] AND [CDU.DH or CDU.TPH contains a relation $Y(CA_y, TC)$ or $Y(TA_y, TC)$ (where Y can be instantiated with any relation)] AND [PDU.DH or PDU.TPH contains a relation $X(CA_x, TC)$ or $X(TA_x, TC)$ (where X can be any relation)] AND [X is not the same as Y] AND [(CA_y, TA_y are in CDU.DH or CDU.TPH) AND (CA_p, CA_x and TA_x are in PDU.DH or PDU.TPH)] AND [CDU.DH or CDU.TPH contains a

[14] UDU was left out of the IS shown in Figure 6.3 for brevity, and contains ungrounded dialogue units prior to PDU. See Chapter 2 or Matheson *et al.* (2000) and Poesio and Traum (1998) for more information on the IS model and fi elds involved.

relation $Z(CA_y, CA_x)$ or $Z(TA_y, TA_x)$] THEN update CDU.DH with CA_z: $concession(X, Y, Z, TC)$

(b) Else if CDU.DH also contains [either an $agree(B_i, CA_p)$, $accept(B_i, CA_p)$ or $acknowledge(B_i, CA_p)$ CA] AND [CDU.DH or CDU.TPH contains a relation $Y(CA_y, TC)$ or $Y(TA_y, TC)$ (where Y can be instantiated with any relation)] AND [PDU.DH or PDU.TPH contains a relation $X(CA_x, TC)$ or $X(TA_x, TC)$ (where X can be any relation)] AND [X is not the same as Y] AND [(CA_y, TA_y are in CDU.DH or CDU.TPH) AND (CA_p, CA_x and TA_x are in PDU.DH or PDU.TPH)] THEN update CDU.DH with CA_z: $concession(X, Y, TC)$

Discussion

As stated previously, the procedure simply assumes that any pair of relations is concessive provided there is minimally acknowledgement of the previous turn. Notice that we do not specify what needs to be agreed with in PDU; as long as anything in PDU is agreed with or at least acknowledged, that is sufficient. Also we did not specify what the TC was in this procedure; the procedure simply specifies that the same TC should be part of both relations X and Y linking CAs or TA in PDU and CDU with this salient claim. Determining the TC relies on the assumption that there is a salient claim which both turns relate to. From the perspective of generating concessive "but", presumably the speaker identifies the TC in PDU and has an alternative perspective w.r.t. this TC and so generates "but" and communicates concession (this will be addressed a bit more in the last chapter).

Recalling that concession (and all discourse relations) are defeasible, this approach does not exclude the possibility that the decision structure above might incorrectly classify an example as concessive. Rather, we err on the side of incorrectly interpreting than not interpreting at all, very much as speakers do in dialogue, since this enables a response, whether corrective or not. In other words, if the procedure incorrectly predicts concession, and then a subsequent response made by the hearer of the concession indicates this concession, which the attributed speaker denies, this is perfectly reasonable. The attributed speaker can simply say "But I never meant Y, and anyway I disagree with what you said earlier (PDU)". Even if we did not overgenerate

interpretations, it is still the case that speakers can deny discourse relations, since the hearer might not have interpreted the previous turn correctly. Consider the example below:

(6.19)

A1: <Didn't you like Ghost World?> p
TC = you liked Ghost World; $X = implicit\_answer(p, question(TC))$

B2: Yes, but <it didn't follow the original graphic novel very closely.> p
$Y = answer(p, question(TC))$; $Z = criticise(p, X)$; $concession(X, [Y, Z], TC)$

A3: How can you say that and still say that you like the film?

B4: Just because the film followed a different story than the novel doesn't mean I didn't also like the film.

B4′: Well, I think the film would have been better if it had followed the novel more closely.

Here A3 directly questions the concession relation interpreted from B2's response. B4 then denies this relation, indicating that she was not criticising the film, just noting a difference between film and novel. B4′ does not deny the concession, and indeed makes her claim more clearly, indicating why she thinks the film was not as good as it could have been.

One last point to note is that in cases in which the scheme in Figure 6.3 holds (with Z relating X and Y), both Y and Z are included in the second slot of the concession relation. This is because both Y and Z together distinguish these cases as concessive.

6.4 Summary

We presented two general schemes characterising what is communicated in cross-turn concession. In the former case (Figure 6.1) we have two turns containing $P1$ and $P2$ respectively, that are both related in different ways to a contextually salient claim, the TC. These relations are $X(P1, TC)$ and $Y(P2, TC)$. The second scheme (Figure 6.3) additionally involves the relation $Z(P2, P1)$ which relates the two propositions directly. We then encountered a non-exhaustive list of constellations (i.e., pairings) of X and Y predicates that corresponded to cross-turn concession in a range of different situations. In all these cases, the speaker of the concessive "but" turn both (1) agreed, accepted,

or minimally acknowledged the former turn ($P1$) while (2) presenting a preferred alternative argument w.r.t. the state described in the TC. Identifying this TC and the relations X and Y enable interpretation of implicitly communicated information that gives useful information on speakers' stance on a relevant topic.

We presented a systematic approach which explored the range of concessive responses to commands, questions and assertions, also accounting for cases in which an answer to a question is responded to concessively. The idea that SAs are wrappers to propositions results in a treatment that does not depend on SAs for the basis of classification, but rather depends on the acknowledgement of something communicated in PDU and the presence of a novel stance w.r.t. the TC to distinguish concession. This bypasses the problem that indirect SAs pose, for example, a question that actually issues a command, e.g., "Can you pass the salt?" now only depends on this agreement and alternate perspective w.r.t. a salient claim for concessive interpretation.

As seen in Example 6.19, the motivation behind this work is to enable hearers to respond to inferred concessive relations, which can then be denied (if disagreed with) by the speaker the relation is attributed to. We cannot claim to be able to interpret defeasible relations correctly always, since this depends on the concessive speaker's mental model and attitude. What we can do is predict relations where they seem reasonable, bearing in mind that the speaker the relation is attributed to can always deny the relation, just as for DofE.

Chapter 7

Correction

In this chapter we will make the claim that cross-speaker "but" can be used to signal correction of something in the previous turn and present a treatment of correction signalled by cross-speaker "but". We argue that correction differs from concession and DofE in that it does not involve inferring relations between the turns themselves or between the turns and a TC, but rather involves disagreement, denial or rejection of something in the previous turn and either an explanation for why this is disagreed with or the presentation of a replacement.

We will start by describing the types of corrections "but" can communicate by focusing on the Speech Act (SA) communicated in the previous turn and address the ways in which "but" can correct what is communicated. We adhere to the idea put forth in the previous chapter that SAs act as wrappers around propositions (Austin, 1962) and aim for a central definition of correction to arise from the analysis. To this end we will focus on previous turns communicating assertions, questions, commands and answers to questions and we will address whether "but" corrects the proposition, the direct SA or the discourse relation communicated in the previous turn. We will briefly investigate similar relations like denial, and explore how one might distinguish denial from correction in cases like the ones presented here.

After presenting a typology of the situations "but" can correct, we will address how these corrections can be modelled in the IS model of dialogue. In the last chapter we will address how modelling "but" can facilitate generation of responses, and we will show how we can deliberate over responses to corrections signalled by "but" given the

model presented in this chapter.

7.1 Correcting Assertions

Assertions can involve a large range of propositions and can include propositions wherein the clauses are related via local discourse relations that only hold between elements in the turn, e.g., assertions of causal relations, temporal ones, exemplification, etc. For example, speakers can assert that event X happened before event Y, or that something caused something else to happen, etc. Assertions can also be related to other turns in the preceding dialogue, or to information inferred from the context (including preceding discourse). For example, speakers can assert something as an example of some prior rule expressed in the dialogue, or as a cause for some prior event, etc. What we will investigate here is how cross-speaker "but" can correct these assertions. Does it correct what is actually asserted, or does it correct the relations asserted or inferred from interpreting the role of the assertion? The example below involves A expressing a reason why chairs have four legs, which is a local causal relation.

(7.1) A: Chairs have four legs for stability.
B: But three-legged chairs can be equally stable, they just need to be well-designed.

Often asserted propositions relate to the preceding discourse, e.g., assertions that serve as answers to questions, as illustrated in the next section. An important observation here is that correction of assertions involves some relation asserted in the preceding turn, or the relation inferred by the corrector between the assertion and the preceding discourse or simply within the assertion itself (as above). Simply asserted material can also be corrected, as in the example above and the following two examples, where the assertions themselves (not relations within the assertion, as in the example above) are being corrected:

(7.2) A: Dogs are trustworthy animals.
B: (Oh,) But they just appear that way because they are stupid.
B′: No they're not, they're just too stupid to be devious.

(7.3) A: The train from Brisbane gets in at 11 o'clock.
B: (Oh,) But that's the Perth train! The Brisbane train only gets in at 1.
B′: No it doesn't, you're thinking of the Perth train.

Notice how both these examples are more definitively negated given a "no" as in B′. Indeed the "but" does not easily correct assertions since it logically communicates conjunction with the implicature of contrast/opposition. These "but" corrections of assertions often seem more understandable with the "Oh" beginning the turn, signalling surprise at the previous turn and indicating understanding and possibly also partial acceptance of the assertion, and disagreeing with either its content or role in the discourse. Disagreeing with an assertion's argumentative stance, usage, or role in the discourse (given discourse history) is similar to corrections of discourse relations which are discussed in the next few sections. It might be the case that "but" corrections of assertions are somewhat softened from a direct negation, but we will not argue this point further without more evidence.

7.1.1 Modelling Corrections of Assertions

In both TOD and NTOD cases, the assertions are denied and then an alternative reason for the manifested behaviour (trustworthiness in Example 7.2 and arriving at 11 o'clock in Example 7.3) is provided. We will sketch out the necessary considerations for generating correction in such cases. The correction procedure will operate on the IS and communicate with theorem-provers capable of determining potential conflicts between beliefs (this will be discussed more below). The abbreviations of fields (e.g., CDU) in the IS are discussed earlier in the thesis and are presented in Poesio and Traum (1998) and Matheson *et al.* (2000).

1. If PDU.DH contains an $assert(speaker[CA_j], X)$ and CDU.DH contains a "but" utterance (i.e., CA_z: $assert[but(Z)]$), then

   (a) Compare X for all assertions in PDU.DH with the CDU speaker's beliefs and expectations in Private Beliefs, his task-related expectations in Task Beliefs and his assertions in UDU.DH and determine via a theorem-prover (or planner for TOD) whether there are conflicts, differences in beliefs, etc. If so, store these conflicts on stack R where each layer of R looks like $[X, Y]$, where X is what is asserted by the previous speaker that is disagreed with

and Y is the current speaker's contradicting/conflicting belief, expectation, etc.

(b) Deliberate over the contents of R and create a list L containing the conflicts which contains the elements of R that were selected to be expressed and possibly other CAs added by the deliberation module.

(c) Repeat the following for all elements in L:

i. Update CDU.DH with CA_k: $disagree(speaker[CDU], L[X])$, CA_l: $correct(speaker[CDU], L[X])$ and CA_m: $assert(speaker[CDU], L[Y])$ for the current element of L.

ii. Remove $[X, Y]$ from L.

iii. Increment k, l, m by 1 each.

Notice that we are only addressing "but" cued corrections here, hence the update procedure is only called if a "but" cue fronts the utterance in CDU.

Comparing X with beliefs, expectations and planning goals, expectations, and previously asserted information in order to determine if the current speaker has conflicting beliefs (even partially conflicting ones) is not trivial, but we will assume that a theorem-prover and planner can handle this[1]. The CA interpreter simply parses utterances into SAs and in the case of questions, into question types following Graesser's taxonomy (Graesser *et al.*, 1992) and then determines whether the following turn's assertion is an answer based on whether it fills the slot or type of information requested by the question. There is no onus upon the CA interpreter to distinguish the correct answer. The TA interpreter interprets Task Actions in TOD, as was described in Chapter 6, where we saw that the TA interpreter takes the current CAs and by communicating with the planner, the TA interpreter resolves referents in the CA to planning operators in the task-plan based on the current stage in the task-plan and updates the TPH with these hypothesised TAs, analogously to the way in which the CA interpreter updates the DH with hypothesised CAs.

[1] The fi elds in the IS are: Ungrounded, Previous and Current Dialogue Unit, (UDU, PDU and CDU); Dialogue History (DH); Conversational Act (CA); CONDitions; Socially Committed Proposition, (scp). See Kreutel and Matheson (2001a) for more information.

In the case of corrections, determining whether there are potentially conflicting beliefs (as mentioned above in the correction procedure) requires theorem-proving and possibly also running planning operators through the planner for TOD. We do not address here how determining conflicts is achieved, since that is beyond the scope of this work, simply assuming that our correction procedure sends the CA in PDU to a theorem-prover which infers beliefs communicated in the CA and then determines conflicts between these new beliefs and beliefs already in the speaker's private beliefs and passes back a stack of these conflicting beliefs to the procedure. In the case of TOD, TAs updated for PDU in the TPH by the TA interpreter get sent to the planner to determine if they are inconsistent with the speaker's task-plan. Inconsistencies are likewise returned in a stack to the correction procedure.

Notice that this interpretation relies on determining the source of conflict itself in the speaker's mental model, which was not really the case for concession, since we assumed there that the concession procedure operated on speakers' CAs. For DofE, we had to check if the speaker of the "but" had such a denied belief in his beliefs. Of course it would be simpler to assume that if both a "but" assertion and a "disagree" CA with something in PDU appear in CDU, we should also update with a correct CA. This procedure does not assume that a disagree CA has been updated. If we had a disagree CA already in the IS, interpretation would not require the theorem-proving it would otherwise require. However in order to determine that the speaker disagrees, theorem-proving would be necessary.

The last point to note is that while we can determine how to find out if there are conflicting beliefs (e.g., via a theorem prover), we cannot determine which of these beliefs should be conveyed. We rely on a deliberation module which considers all these conflicting beliefs and other information before determining what should be generated in response.

7.1.2 Correcting Answers

Answers to questions are SAs responding to preceding SAs (relational SAs, following Poesio and Traum (1998), and "but" can correct the answerhood relation itself, as presented in the correction conversation given in Asher (1998), which is simplified

below:

(7.4)
A: Why did (John get sent to jail)? *r*
B: (He was caught embezzling funds from the pension plan). *p*
C: Yes, (BUT (he went to jail) *r* because (he was convicted of tax evasion) *q*1) *q*2

This case is particularly interesting because it can be neatly analysed via the treatment of concession presented in the previous chapter. The observation that certain types of corrections bear close resemblance to cross-turn concession leads to two objectives. Firstly, this implies that the treatment of cross-turn concession will probably be easily adapted for these kinds of concession (and vice versa). Secondly, their similarity requires clearly distinguishing the two. We follow the semantic representation scheme for concession presented in Grote *et al.* (1995), modified to address cross-turn concession, where one speaker's turn argues in favour of some contextually available claim or Tertium Comparationis, (TC, following Lagerwerf (1998)), while the other speaker's turn argues against this claim or TC. Grote et al's model of concession posits (1) that some situation A holds, implying the expectation that C holds, and (2) that another situation B also holds which implies that C does not hold.

We address here the connection between this concessive interpretation and question-answering situations like Example 7.4 above, where the question provides the contextually available TC (phrased as the claim that John got sent to jail rather than as a question), and B gives an answer which C accepts as an assertion, but rejects as an answer to the question in A. The similarity with concession (as defined in Chapter 6) arises from C's acceptance of B's assertion (p) and subsequent assertion of an alternative antecedent that answers the question. This similarity is due to B and C providing alternative answers to A's question, although notice that neither either implicates or denies the TC. That is, correction differs from concession because it does not involve implicated expectations. C rejects B's assertion as an answer to A's question (i.e., $\neg[p > r]$). We will define correction here as both (1) a rejection of B's utterance as an answer to A, and (2) the assertion of an alternative answer to A.

However correction of an assertion following a question does not necessarily mean that both assertions (if the correction takes the form of an assertion) are answers, as illustrated in B3 in the following example:

(7.5)
B1: Why are cats selfi sh?
A2: Cats are independent creatures.
B3: But that doesn't mean they have to be selfi sh.
B3′: But then why do they only stick around if you feed them?

Distinguishing between concession, DofE and correction

Notice that Grote et al.'s scheme distinguishes both DofE and concession as well, in these cases involving implicated expectations. For concession, the speakers argue for and against a claim, while for DofE, the second speaker simply denies the claim. Using their terminology (i.e., A and B for the situations and C for the claim), concession involves $A > C$ and $B > \neg C$ while DofE involves $A > C$ and $\neg C$ (or $B = \neg C$) in the two turns.

This is in contrast to the Lascarides and Asher (2002) analysis of the monologue example "[John bought an apartment] but [he rented it]". They argue here that the clauses are related by both Narration and Contrast relations, where the correct temporal effects arise via the Narration relation. However their Contrast relation does not capture the idea that the first clause implicates the expectation that one buys an apartment to live in it, which is denied by the second clause. I.e., it does not address argumentative aspects of "but" (i.e., concessive interpretation) or the denied expectations presupposed by "but". This example could certainly be analysed as concessive, with the TC "John moved into the apartment", and the first clause arguing in favour while the second one argues against. In fact, it is difficult to see how the definition of Contrast in Asher (1993) can be extended to address situations like this when there is very little in common between the clauses in order to draw the distinction necessary for Contrast. Indeed his Contrast relation seems closer to the definition of semantic opposition (SO) put forth in Lakoff (1971). For example, consider the example below, which does not involve isomorphic structures and is much harder to consider as SO, with p, q, r shown to illustrate how they map onto the basic scheme we argue for:

(7.6)
A: Bob intended to go. p
B: But he had visitors. q
r = Bob went.

Following Asher (1993), this example does not meet the criteria of Contrast, and yet SDRT does not address concession, which seems the best analysis of both this sort

of example and the embezzlement example, with r representing the TC, since in both cases, $p > r$ and $q > \neg r$. Consider also the example taken from Asher and Lascarides (1998b) addressing question-answering in more task-oriented situations:

(7.7)
A: How can I get to the treasure? r
B: It's at the secret valley. p
A: But I don't know how to get there. q

A responds here to the implicit assumption in B that A knows how to get to the secret valley, and denies that this assumption holds. What this example gets at is how the nature of presuppositions launched by definite descriptions in dialogue situations is extended, so that "the secret valley" does not simply presuppose that a secret valley exists, as in monologue, but that it exists and *also* that the addressee knows (1) that the secret valley exists and (2) knows which valley the speaker is referring to in case of multiple secret valleys. I.e., in TOD, B needs to provide other information which are preconditions that A needs to get to the treasure, e.g., knowing both where the treasure is and how to get there. The assertion that the treasure is at the secret valley presupposes that the addressee knows where the secret valley is, or else it would not be sufficient for B to simply indicate that the treasure is at the secret valley. This can also be analysed as following Grice's Conversational Maxims (Quantity).

Resolving Example 7.4

Focusing in greater depth on Example 7.4, we will try to get at precisely what enables us to resolve this example. Our goal is to resolve it into first-order logical variables that we can reason with via the p,q,r scheme described above for concession, so that we can determine what is being communicated in these corrections of answers. Considering the embezzlement example, the *why*-question enables interpretation of B's turn as indicating a *reason* (and therefore an answer) to the situation for which A asks an explanation.

We will take r to be the situation queried about in A, so r is "John got sent to jail". We will also advocate replacing $p > r$ with $\mathit{reason}(p,r)$ for B's turn, since we know that this more specific relation holds by interpreting the assertion as an answer to the *why*-question. Of course, for other types of questions (e.g., *how*-questions), this relationship will not be appropriate and we will simply have the defeasible implication $p > r$. Also, this analysis will not work in non-question-answering situations where

we cannot interpret B's turn as an answer w.r.t. a question. The question sets up an expectation of an answer, which constrains interpretation. Furthermore, we will assume that CAs like assertions, answers, questions ("ask" CAs in the figure below), etc. are determined via prior interpretation procedures, e.g., by a CA interpreter.

Given these assumptions, we can see (as marked in the embezzlement example) that A communicates *r*, and B communicates *p* and implicitly (via interpretation) *reason*(r, p), assuming that *reason* is a binary predicate expecting a situation (*r*) and reason (*p*) for the situation. We interpret *p* (given the CA interpreter) as (1) an answer expressing (2) the reason for *r* and also, more basically (3), as an assertion of the situation described in B. Then C's "yes" is interpreted as accepting B's assertion, and the "but" clause ("but he went to jail because he was convicted of tax evasion", *q*2 in the example) indicates an alternative answer w.r.t. B's turn, since the first part of *q*2, "he went to jail" is *r* restated, and is explained (cued by "because") by *q*1 ("he was convicted of tax evasion"), which is expressed as a reason for *r*, so we get *reason*$(r, q1)$. Since we know that for a *why-r* question, anything which involves *reason*(r, X) is an answer to the question, we can interpret this as an alternative answer to A's question. The "but" indicates that C's utterance is somehow contradictory to B's utterance, so we interpret this alternative answer as a correction of B, i.e., both (1) a rejection of B's utterance as an answer to A, and (2) the assertion of an alternative answer to A, which we will use as a rough definition for corrections like these.

Notice that the dialogue in the embezzlement example would not be coherent without the "but" in C's turn. I.e., if the "but" were omitted, we would assume that C failed to interpret B's turn as an answer to A. So the "but" indicates a non-defeasible sense of contrast here, and we can argue that when "but" signals correction, it introduces non-defeasible contrastive information.

Without the question in A, i.e., in a dialogue consisting solely of B's and C's turns, we would not be able to interpret B as answering A. In fact, we would not be able to interpret B as providing a reason for *r*, since *r* itself is only introduced in C's turn. However, once C's turn is uttered, then we can interpret *reason*$(r, q1)$ as before.

7.1.2.1 Modelling Correction of Answers to Why-Questions

The goal of modelling corrections to answers of *why*-questions in the IS model of dialogue is (1) to distinguish these corrections from other similar "but" concessions and (2) to update the IS with precisely what has been communicated in order to enable more accurate response generation by the system in the subsequent turn (assuming the system plays the role of speaker B). We focus here on modelling what the speaker of the correction (C) is communicating, rather than the interpretation of the prior turn as an answer, and make use of the CA interpreter to determine whether the current and previous turns contain answers to a preceding *why*-question. (Recall that there is no onus upon the CA interpreter to check if the answer is correct; it only determines answerhood by checking if the answer fills the queried slot-value in the question or matches one of Graesser's types (Graesser *et al.*, 1992). The CA interpreter simply takes utterances and maps them into their corresponding CAs; it can also interpret whether a given CA is an answer to the previous turn's question. We also assume that the question is answered and then that the answer is corrected immediately with no intermediate turns. The goal here is to specify the necessary conditions for updating the IS with a correction SA.

1. After a new turn, pass this utterance to the CA interpreter. If [UDU.DH contains an $ask(why[X])$ SA in CA_i] AND [PDU.DH has a CA_j that is of the form $assert(Y)$] AND [CDU.DH has a CA_z of the form $assert(but[Z])$] then

   (a) If the $CA\_interpreter(IS, CA_j, R)$ returns $R = answer(speaker(CA_j), CA_i)$[2] then

      i. If CDU.DH contains a CA_k with the SA $assert(Z)$ and if the $CA\_interpreter(IS, CA_k, R2)$ returns $R2 = answer(speaker(CA_k), CA_i)$, then add to CDU.DH CA_m: $reject(speaker(CA_k), R)$ and CA_n: $correct(speaker(CA_k), R)$

The update procedure interprets whether correction occurs and then updates the CDU with correct, reject and possibly accept CAs. Notice that we cannot add the

[2] By *answer* here we make no claims that this is the correct answer from any perspective of the dialogue participants.

C: Yes, but he went to jail because he was convicted of tax evasion.

$$\left[\begin{array}{ll}
\text{UDU} & \left[\text{DH [CA1: } ask(A,(B,C),why(sent(John,to(jail))))\right] \right] \\
\text{PDU} & \left[\text{DH} \left[\begin{array}{l}\text{CA2: } assert(B,caught(John,embezzling(funds,from(pension_plan)))) \\ \text{CA3: } answer(B,CA1)\end{array}\right]\right] \\
\text{CDU} & \left[\begin{array}{ll}
\text{DH} & \left[\begin{array}{l}\text{CA4: } accept(C,CA2) \\ \text{CA5: } assert(C,reason[sent(John,to(jail)),convicted(John,tax_evasion)]) \\ \text{CA6: } answer(C,CA1) \\ \text{CA7: } reject(C,CA3) \\ \text{CA8: } correct(C,CA3)\end{array}\right] \\
\text{COND} & [accept(B,CA6) \rightarrow scp(B,CA7)]
\end{array}\right]
\end{array}\right]$$

Figure 7.1: IS for Example 7.4

social commitment that if the hearer of the correction accepts the correction then he must accept the alternative answer, because he could accept the correction for other reasons than the alternative answer.

7.1.3 Correcting Answers to Other Types of Questions

This leads us to investigate whether corrected answers to other types of questions can also be modelled like the *why* question above. We take Graesser's taxonomy of inquiries (Graesser *et al.*, 1992) as a basic set of question types and omit the categories in Graesser's taxonomy which involve single-valued (e.g., slot-filling) answers and prefer those which tend to require answers which attribute some predicate to a subject (so we have sentential rather than phrasal answers); see Table 7.1. This leaves us with questions involving comparison (e.g., how is X different from Y?), definition (e.g., what does X mean?), possibly interpretation questions, and feature specification. We also take his question categories involving causal antecedents and consequents (e.g., what caused X? and what are the consequences of X?), goal orientation (e.g., what are the motives/goals behind an agent's actions?), instrumental/procedural questions (e.g., how does an agent achieve a goal?), and expectational questions (e.g., why did some expected event not occur?).

Since several categories in his taxonomy have questions beginning with "what", some of which can also have *why*-questions, and the remaining three categories have

Table 7.1: Question-Types from Graesser's Taxonomy which we Account for

Question	Abstract Specification	Question Word	Answer Type
Comparison	How is X similar/different to/from Y?	How	Assertion
Definition	What (category/properties) does X have?	What	Assertion
Interpretation	How is an event interpreted/summarised	What/How/etc.	Assertion
Feature Specification	What value/attribute does feature X have?	What	Assertion[4]
Causal Antecedents	What caused event X to occur?	Why/How	Assertion
Causal Consequents	What are the consequences of an event/state?	What	Assertion
Goal Orientation	What are an agent's goals/motives?	Why/What	Assertion
Instrumental	How (plan) does an agent accomplish a goal?	How	Assertion
Expectational	Why did some expected event not occur?	Why	Assertion

how-questions, we will assume that we have the appropriate machinery to resolve several question-types.[3] The benefit of using the taxonomy is that it provides us with useful clues about the nature of the answer, supposing the answerer to be honest and helpful (following Gricean reasoning). We also assume (throughout this thesis) that the IS contains the relevant private beliefs etc. required to determine the hearer/corrector's stance w.r.t. the answer.

7.1.3.1 Modelling Correction of Answers to Other Question-Types

Now if we assume that we can classify a question into these categories, we can simply incorporate these categories into the framework presented in the previous section for modelling correction in the IS update model of dialogue. Consider the modified procedure below:

1. [If CDU.DH has a CA_z of the form $assert(but[Z])$] AND [if UDU.DH contains $question(speaker[UDU], W, X)$ in CA_i, where X corresponds to one of

[3] We should point out that this is not such a far-fetched assumption; Graesser's taxonomy is in active use by the Question Answering community, and the categories are presumably resolvable.

Graesser's categories (Table 7.1)[5]] AND [PDU.DH has a CA_j that is of the form $assert(Y)$] then

(a) If the $CA\_interpreter(IS, CA_j, R)$ returns $R = answer(speaker(CA_j), CA_i)$ then

i. If CDU.DH contains a CA_k with $assert(but[Z])$ and if the $CA\_interpreter(IS, Z, R2)$ returns $R2 = answer(speaker(CA_k), CA_i)$, then

A. If CDU.DH contains a CA_l with $accept(speaker(CA_k), CA_j)$ then add to CDU.DH CA_m: $reject(speaker(CA_k), R)$ and CA_n: $correct(speaker(CA_k), R)$

B. else add to CDU.DH CA_m: $correct(speaker(CA_k), R)$ and CA_n: $implicit\_disagree(speaker(CA_k), CA_j)$

Given the triggering construction of "but" in the current turn and a question two turns back followed by an answer in the previous turn, the interpretation procedure determines whether this answer is being corrected, and if so, updates the IS with correction and rejection of the previous answer, and, if applicable, implicit disagreement with the assertion itself in the previous turn. Notice that we do not know in the latter case whether the corrector disagrees with the assertional content of CA_j or its answerhood; we assume that CA_j's answerhood is being corrected since the corrector introduces an alternative answer, however, since she can disagree with either CA_j's assertional content or answerhood, we simply update that she implicitly disagrees (since she does not give explicit disagreement) with CA_j itself, leaving what exactly is disagreed with ambiguous in the update.

Of course, what is particularly interesting given different question types (as opposed to the more simplistic *why*-question case we saw earlier) is that provided we have a CA interpreter that can parse questions into the Graesser types discussed above, we now have much more specific information about both questions and answers be-

[5]It is not really necessary as things stand in the algorithm at present to refer to Graesser's categories here since they are not utilised at all. They were included here since they might eventually be of help given a CA interpreter capable of determining which category (according to Graesser) a question falls into and then resolving how the answers fi t into this category.

ing corrected, and more specifically, information on what exactly the correction hinges upon. In the subsequent sections of this chapter, we will try to extend this basic procedure for other cases in hopes of gaining a basic update strategy for correction.

7.2 Correcting Discourse Relations and Implicit Communication

Implicitly communicated information like discourse relations, denied expectations, inferences, and defeasible rules can all also be corrected, so we will explore these cases briefly here.

Recall Example 8.3 (A: "Chairs have four legs for stability.") where A's assertion communicates a reason for the stability of chairs, which is then refuted by B ("But three-legged chairs can be equally stable, they just need to be well-designed."). Here B refutes the *reason* relation communicated by A by directly refuting the inference that A communicates that $four\_legs(chair) > stable(chair)$. The embezzlement example showed that C could agree with the assertion B made but disagree with the inferred answerhood function of the assertion (which makes the assertion coherent given the preceding question). The idea here is to model corrections of other sorts of implicit information, like discourse relations in Example 8.3.

7.2.1 Modelling Corrections of Some Implicit Information

We describe here a mechanism for determining whether DofE, concession, alternatives, planning proposals or any additional expectations, rules and beliefs should be corrected. Notice that this is a procedure for deliberating over a prior turn's implicitly communicated information (e.g., discourse relations), and enables the hearer to respond appropriately depending on her own beliefs w.r.t. this implicit information.

1. If CDU.DH contains CA_z: $assert[but(Z)]$, do the following for all updates of the form $X(Arg)$:

   (a) If $X = dofe$, $Arg = [Speaker, Expectation]$, pass the CA $dofe(Arg)$ and a pointer to the IS to a theorem-prover which compares Expectation with

(in the IS) all of speaker CDU's beliefs in Private Beliefs; pass the TA update corresponding to this CA to the planner to determine if there are any planning conflicts. If there are conflicts, differences in beliefs, etc, these are returned in a stack (R) from the theorem-prover. Likewise if there is conflict with the denial of the Expectation in the IS, push this conflicting information onto stack R.

(b) If $X = concession$, $Arg = [CA_j, CA_i, TC]$ where CA_j is in PDU.DH, pass $concession(Arg)$ and a pointer to the IS to the theorem-prover which compares CA_j, and TC with (in the IS) all beliefs in Private Beliefs and all assertions (in DH) and likewise pass the TA equivalent of the CA to the planner to determine planning conflicts. If there are conflicts, differences in beliefs, etc, these are returned in a stack (R) from the theorem-prover/planner. Similarly compare the rules $CA_j > TC$ (or $CA_j > \neg TC$ if that is predicted by the CA interpreter) with beliefs, expectations, planning rules and assertions and push any arising conflicts onto R.

(c) If $X = alternatives$, $Arg = [CA_j, CA_i]$ where CA_j is in PDU.DH, pass $alternatives(Arg)$ and a pointer to the IS to the CA interpreter which compares this with (in the IS) all beliefs in Private Beliefs and all assertions (in DH) and task actions (in TPH) communicated by CDU speaker; if CA_i and CA_j are not alternatives in her beliefs, expectations, etc., this is returned in a stack (R) from the theorem-prover.

(d) If $X = propose\_plan$, $Arg = [planning\_operator/s]$, pass $propose\_plan(Arg)$ and a pointer to the IS to the planner which compares Arg with (in the IS) the speaker's task-plan and task beliefs. If there are conflicts, differences in plans, goals, etc, these are returned in a stack (R) from the planner.

2. Deliberate over the contents of R and create a list L containing the conflicts which contains the elements of R that were selected to be expressed and possibly other CAs added by the deliberation module.

3. For list L (each element of L looks like $[X,Y]$, where X is the assertion disagreed with and Y is the contradicting/conflicting belief, expectation, etc.), repeat the following until L is empty:

   (a) Update CDU.DH with CA_k: $disagree(speaker[CA_k], L[X])$, CA_l: $correct(speaker[CA_k], L[X])$ and CA_m: $assert(speaker[CA_k], L[Y])$ for the first element of L and then remove $[X,Y]$ from L. Increment k,l,m by 1 each.

This procedure updates the IS with the appropriate information indicating correction (if it is predicted by the procedure) and facilitates deliberation of appropriate responses. As in the case of assertions, the major work is done behind the scenes by the theorem-prover and planner (for TOD), and we simply assume for now that given the relevant information, the theorem-prover can determine whether there is any conflict between speakers' beliefs given what was just uttered. If so, the current speaker's IS is updated with CAs indicating that she disagrees (and with what), that she makes a correction (and of what), and that she asserts the alternative information she believes to be true. We pass both discourse relation and arguments and IS (or in practice, a pointer to the IS) to the theorem-prover so that it has a sense of context, e.g., that an expectation was denied, and what that expectation was. Depending on which discourse relation is communicated, the theorem-prover makes the appropriate inferences and determines whether the current speaker has any contradicting information to these expectations and relations. If so, she can respond appropriately, indicating that she disagrees.

As in all the algorithms presented in this chapter, we generate the disagreement necessary to invoke correction here. To start with we can always assume that given disagree or reject (with something in the prior turn), correct and assert (of an alternative) SAs in a "but"-cued turn, updating with correction is a valid (though defeasible) interpretation. This is illustrated below:

1. If CDU.DH contains CA_z: $assert[but(Z)]$ (and assuming last speaker is B), and if CA_j: $disagree(B, CA_i)$ (in CDU.DH), where CA_i was uttered previously (in PDU.DH) by A, and CA_b: $assert(B,Y)$ in CDU.DH, repeat the following for all CAs of the form $X(Arg)$ in PDU.DH (the preceding turn):

(a) If $X = dofe$, $Arg = [A, Expectation]$, pass $dofe(Arg)$ and the IS to the theorem-prover which compares Expectation with Y (in CA_b). If they conflict, update CDU.DH with CA_m: $correct(B, Expectation)$

(b) If $X = dofe$, $Arg = [A, Expectation]$, pass $dofe(Arg)$ and the IS to the theorem-prover which compares denial of the Expectation with Y (in CA_b). If they conflict, update CDU.DH with CA_m:
$correct(B, dofe[Arg])$

(c) If $X = concession$, $Arg = [CA_j, CA_i, TC]$, pass $concession(Arg)$ and a pointer to the IS to the theorem-prover which compares CA_j with Y (in CA_b). If they conflict, update CDU.DH with CA_m: $correct(B, CA_j)$

(d) If $X = concession$, $Arg = [CA_j, CA_i, TC]$, pass $concession(Arg)$ and a pointer to the IS to the theorem-prover which compares TC with Y (in CA_b). If they conflict, update CDU.DH with CA_m: $correct(B, TC)$

(e) If $X = alternatives$, $Arg = [CA_n, CA_i]$ where CA_n is uttered by B in some preceding turn or in the current turn), pass $alternatives(Arg)$ and a pointer to the IS to the CA interpreter which compares CA_n with CA_b. If they conflict, update CDU.DH with CA_m: $correct(B, CA_n)$

(f) If $X = alternatives$, $Arg = [CA_n, CA_i]$ (where CA_n is uttered by B in some preceding turn (or in the current turn), pass $alternatives(Arg)$ and a pointer to the IS to the theorem-prover which compares
$alternatives(Arg)$ with CA_b. If they conflict, update CDU.DH with CA_m:
$correct(B, alternatives[Arg])$

(g) If $X = propose\_plan$, $Arg = [planning\_operator/s]$, pass $propose\_plan(Arg)$ and a pointer to the IS to the planner which compares Arg with Y (in CA_b). If they conflict, update CDU.DH with CA_m:
$correct(B, planning\_operator/s)$

This last approach assumes we already know that B disagrees with something A said before, and infers that if B disagrees with something and provides an alternative to what she disagrees with, then she is correcting this information of A's. Note also that cases a-e above assume a "but" in PDU as well as CDU. This procedure is an

alternative to the former one which assumes that the disagreement and assertion CAs have already been interpreted, and indicates that the procedure presented previously can be broken up into a set of interpretation stages that operate sequentially, where the interpretation of disagreement and assertion are assumed to have already been achieved at the point at which the above procedure is called.

7.3 Corrections and Questions

Questions themselves can be corrected, but what does it mean to correct a question? While there is much work involving the semantics of questions (e.g., Asher and Lascarides (1998b), Ginzburg (1995a), Ginzburg (1995b), Ginzburg (1996)), Asher and Lascarides argue that much of it falls either into the realm of dialogue planning or formal semantics, and neither type of approach bridges the gap in order to explain examples like the one below:

(7.8)
A: How do I install the assembly?
B: Typically, you put the assembly on before tightening the screws on the assembly.
A: But I tried that and it was too difficult.
B: So let's tighten the screws on the assembly first.

Here B's answer introduces domain knowledge that A deems irrelevant for achieving the desired action. Asher and Lascarides argue that this example requires using compositional semantics to learn different ways of performing an action, thereby posing a challenge to both planning and semantic approaches. They propose combining the approaches, so that B's alternative answer arises by combining semantics and domain plans in his knowledge base, which is similar to what we advocate, namely involving different aspects of speakers' mental states in the interpretation process.

In Example 7.7 discussed earlier, we saw how our approach would indicate that there is something strange if A responds to B's answer (that the treasure is at the secret valley) by saying "But I know how to get there", since "but" indicates that something is in contrast, as Asher and Lascarides point out in discussion on Example 7.8 above if A were to say instead "But that's fine". Of course this response could also be accommodated by assuming that A is surprised at the meta-level knowledge that she

knows where the secret valley is (or knows how to put on the assembly in 7.8 above). Asher and Lascarides argue that this response (i.e., "But that's fine") is incoherent, while we argue that there is a context in which it is coherent, namely one in which A is surprised that she knows where the secret valley is or how to put on the assembly.

Their approach follows from Ginzburg (1996) in his multiple notions of answerhood, namely that (1) the information fully resolves the question, defined in terms of the interpreter's goal and mental state, and (2), that the information potentially resolves the question. That is, Ginzburg's notions of answerhood rely on context sensitivity and interpreter-specific responses. Ginzburg's analysis identifies a proposition at the centre of the question, e.g., in the question above, the proposition would be "I install the assembly". In this sense, correcting a question itself would be very similar to correcting an assertion[6] except that it would involve adjustments to the answerer's obligations.

Also, different question types all expect specific answers, e.g., *why, how, what, when,* and *where* questions; in the case of "why X", the answer is usually a reason for X, where X is some proposition describing a state of events or situation. "How X" expects an answer that provides a manner in which X might be accomplished, or a way to perform/achieve (X), where X is again a proposition containing a state of events. "What X" questions are less specific and X is often a phrase which is usually the subject of a transitive verb, so that the answer provides the direct object.

Of course these sketches of question answering are very rough, and there is far more extensive work on the subject. The point here is that if we again adopt Graesser's taxonomy of inquiries discussed earlier in the section on answers, we get much more specific requirements for answerhood (see Table 7.1). In any case, our goal is to see how this information can inform a more specific characterisation of the types of corrections that are coherent given the preceding question's context. Given such specific information about what a question addresses and what sort of answer it expects, it becomes less difficult to see how the questions in the table can be corrected. Although one difference from correcting answers is that correcting questions points out how the question itself is invalid/incorrect/irrelevant, rather than disagreeing with and providing an alternative answer. Here are some examples corresponding to the first few

[6] Thanks to Colin Matheson for this point.

categories discussed in the table:

(7.9)

Comparison:	But X is the same as Y!
Definition:	But X is undefined!
Interpretation:	But it (the event) never happened!
Feature Spec:	But I already told you!
Causal Ante:	But nothing caused X to occur!

So we notice that the corrections can deny the basis of the question, for example, that a comparison is valid in the first question-type (Comparison) above. The correction can also prove the question invalid, as in the Interpretation case, where it is impossible to interpret an event that never happened. Likewise, corrections can address meta-level issues as in Feature Specification above; here the corrector indicates that the question itself has already been answered. Notice the strong role played by the question category above; in many cases the correction hinges on the validity or relevance of the question category itself.

Another sort of correction of questions involves incorrect assumptions of slot-values in the question. Correcting misassumptions communicated in the question signals a difference in speakers' beliefs, as seen below:

(7.10)

A:	When did you want to fly back from Boston?
B:	But I want to fly back from New York!
B′:	But I want a rail ticket!
B″:	But I don't want to leave Boston at all!

In B a slot-value is corrected, namely place of departure, and then presumably the question becomes valid according to B. B′ indicates that the verb slot-value is incorrect; it should be "travel by train" instead. B″ questions the validity of the question itself, since A asks when B wants to travel and B does not want to travel at all.

Another type of correction involves situations in which the questionee finds something wrong with the central proposition being questioned, as in the example below:

(7.11)

A:	Why are dogs vegetarian?
B:	But they aren't vegetarian!

Here B corrects the proposition at the centre of the question, i.e., "dogs are vegetarian", by denying what is predicated of the subject. This invalidates the question itself. Notice that the question does not ask about this proposition's truth and instead

assumes the truth of this central proposition. The predication in this case is denied; in the following example the property linked to the subject is invalidated:

(7.12) A: Will dogs eat cardboard?
B: But cardboard is inedible.

This example shows how yes/no questions can have their central proposition corrected. Notice that none of the categories in Table 7.1 apply to either of these two cases. In 7.9 we showed how different question types can be corrected which involved showing how the question-type was inappropriate given the proposition which was queried. 7.11 is quite similar to the Definition and Feature Specification categories in Table 7.1, but asks for a reason to answer the question rather than a feature-value or defining category. Both Definition and Feature Specification categories take *what* questions which have as central propositions (for example) “dogs have category Z” and “dogs have attribute Z” respectively. Both these cases involve incomplete propositions rather than complete propositions which are queried as in both Example 7.11 and 7.12 above, and the point here is simply to illustrate that even cases involving complete propositions can be corrected, and in this case simply by showing that the central proposition is somehow invalid or inappropriate. Notice also that both these cases have the general properties of correction, namely (1) showing that something is incorrect and (2) giving a reason why. In these cases, the answers simply state this reason directly, and the criticism is inferred from this.

In the example below, we consider the situation in which the question refers to the preceding turn’s proposition. It is somewhat harder to conceive of discourse relations linking questions to previous discourse directly, although they can sometimes convey requests indirectly as below, or implicitly communicate a discourse relation with the previous turn, as in Example 7.14 (following the one below).

(7.13) B: I’m going for a walk.
A: Will you take Hilda with you?
B: But she’s at school now and I can’t wait.

Here B responds by indicating precondition failure; B wants to go now and Hilda is not with B and therefore cannot come. B answers A here while correcting A’s assumptions of B’s goals. A similar invalidation occurs in the example below. We consider

this correction because a reason is provided for why the question is deemed invalid. In the example below, A suggests a solution to B's problem of powering the circuit, and B corrects A's assumption that the batteries bought yesterday will work with the circuit by introducing information about the circuit's power needs that conflicts with the batteries A suggests using.

(7.14)
B: I need to power the circuit somehow.
A: Will the batteries we bought yesterday work?
B: (Oh,) But we need a 5V battery and the ones we bought yesterday were 1.6V.

7.3.1 Modelling Corrections Involving Questions

So the ways in which questions can be corrected are as follows:

- Correcting an incorrect slot-value (Example 7.10)
- Indicating that the question is not valid
  - Because a necessary criterion/precondition/constraint has not been met (Examples 7.13 and 7.14)
  - Because the question-type itself does not apply to what is being asked about (See 7.9, e.g.: Comparison, Definition, Causal Antecedent)
  - Because it was already asked and so is redundant (E.g., Feature Specification)

The relationship to the prior discourse does not play a significant role in the correction of questions. Rather, we focus on the hearer's mental model to determine whether there is something inappropriate about the question. With these considerations in mind, we sketch below the criteria for interpreting correction, given a "but" turn following a question turn:

1. If CDU.DH contains a turn-initial "but" and PDU.DH contains CA_j: $question(speaker[PDU], X, T)$ (where T is the question's category in Graesser's taxonomy)

(a) If CDU.DH contains CA_k: $disagree(speaker[CDU], part\_of[X])$ and also CA_l: $assert(speaker[CDU], Y)$ and calling the theorem-prpover with $part\_of(X)$ and Y returns that they are alternate values for same attribute or that both share a topic, then add to CDU.DH
CA_m: $correct(speaker[CA_k], part\_of[X])$

(b) If CDU.DH contains CA_k: $disagree(speaker[CDU], X)$ and also CA_l: $assert(speaker[CDU], Y)$ and calling the theorem-prover with X and Y returns that they are alternate values for same attribute or that both share a topic, then add to CDU.DH
CA_m: $correct(speaker[CA_k], X)$

(c) Else if CDU.DH contains CA_k: $disagree(speaker[CDU], CA_j)$ and also CA_l: $assert(speaker[CDU], Y)$ and calling the planner with TA_j (the TA equivalent for CA_j, X and Y returns that Y is a precondition/ constraint-to-be-overcome of X, then add to CDU.DH CA_m:
$correct(speaker[CA_k], CA_j)$

(d) Else if CDU.DH contains CA_k: $disagree(speaker[CDU], CA_j)$ and also CA_l: $assert(speaker[CDU], Y)$ and calling the theorem-prover with CA_j and Y returns that $reason(incompatible[X, T], Y)$, then add to CDU.DH CA_m: $correct(speaker[CA_k], CA_j)$

(e) Else if CDU.DH contains CA_k: $disagree(speaker[CDU], CA_j)$ and also CA_l: $assert(speaker[CDU], Y)$ and CA_j appears in UDU.DH, then add to CDU.DH CA_m: $correct(speaker[CA_k], CA_j)$

The last case in the above procedure catches other sorts of question corrections, where the question can occur further back in the dialogue than the immediately preceding turn. It is possible that this last case may overgenerate corrections, as there is no call to the theorem-prover to determine that Y actually corrects CA_j.

7.4 Corrections and Commands

Plans and interactions between speakers' plans often provide the source of disagreement evident in corrected commands. In the example below, B signals precondition failure (in order to shut the door, it must be open) and corrects A by asserting this.

(7.15)
A: Shut the door.
B: But it's already shut.
B′: But then it'll get too hot; why don't we shut the window instead?
B″: But then it'll get too hot.

B′ communicates an undesirable effect of performing A's commanded action, and proposes an alternative. B″ just communicates the undesirable effect and does not propose an alternative solution, and we do not consider this correction. We will assume that correcting commands like questions and assertions also involves both (1) disagreeing or denying something in the previous turn and (2) proposing an alternative. The question then arises as to why B above is considered a correction, since it does not provide an alternative. We argue that since B shows that A's action is invalid or impossible, it can be considered correction in a way that is very similar to how the various question categories in Example 7.10 were deemed irrelevant or invalid given the subject of questioning. So in both these situations, the assertion part of correction involves asserting what or why the question or command is invalid or irrelevant. The difference between B″ and B above is that B″ only presents an undesirable effect and neither invalidates the command nor rejects and presents an alternative, so it is not deemed as correction. B on the other hand asserts that the commanded action is impossible (rejection/denial/disagreement) and presents an alternative, and so is acceptable as a correction.

Speakers can also propose better alternative actions, as seen below, where the rationale behind A's command is brought into question by the contradictory fact that the flies are getting in through the windows more than the door:

(7.16)
A: Shut the door so the flies don't come in.
B: But the flies are getting in through the windows more; we should probably shut those instead so we can still have some breeze from the door.

Speaker B can also object to the discourse relation inferred from A's assertion. In the example below, B infers that A's commanded action is intended to achieve the goal of keeping the flies out, assuming A correctly interprets B's assertion of the problem that there are too many flies inside.

(7.17)
B: These flies are really getting to me.
A: So shut the door.
B: But they are coming in through the windows.

In this case, what is actually being corrected is the inferred relation between the command and the problem mentioned in the preceding discourse. B is correcting A's assumption that the flies are getting in through the door, and therefore corrects A's solution relation to B's problem, rather than the command A issues by showing that A's solution is irrelevant. This is not correction of a command itself, so we will not discuss it further here.

Of course many of these cross-speaker "but"s indicate disagreement, conflict, denial when responding to a command, as with assertions, questions, answers and implicit information, and like in these cases, we will require that an alternative or explanation is asserted as well.

7.4.1 Modelling Corrections of Commands

We will assume that commands follow questions and assertions in (1) denying the command and (2) either asserting an alternative or showing why the command is not valid or relevant. We will take into account the role of prior discourse (as seen in Example 7.17) to take advantage of the role of the command in achieving the goal, solving the problem, etc. in order to compute how the correction relates to the command and preceding discourse.

The statement of a problem, e.g., B's "These flies are really getting to me" in Example 7.17, is translated into a goal where this effect either does not occur or is mitigated; this is the goal of the second step of the procedure below. In Example 7.17, B communicates an alternative goal, presumably that of shutting the windows, by correcting A's misassumption about the reason for the problem of the flies getting in. The statement of a problem in the CA is translated into a goal via the TA interpreter upon hearing

B's utterance; as mentioned earlier, the TA interpreter resolves referents via the current stage in the task plan and with the help of a planner, and returns hypothesised task actions communicated by a given CA.

The first case in the procedure (a) pertains to the situation in which a plan or goal is proposed and the command attempts to achieve it in a way that is deemed incorrect by the corrector (and speaker of the "but" turn, CDU below) because either there is a constraint between the plan and the commanded action, or the commanded action has a negative effect. Step (b) follows Example 7.17 and corrects the commander's misassumption about what will solve the problem (in this case, she believes that closing the windows will solve the problem). Given a statement of the problem (or situation, Z below), the speaker of the "but" believes that the goal is to avoid Z. Notice that here that the previous turn's command X does not solve Z. Here Y indicates that X does not solve Z, as in Example 7.16B. Step (c) describes situations in which an undesirable effect of the commanded action is pointed out, and an alternative action introduced (see Example 7.15B$'$).

1. If PDU.DH contains CA_j: $command(speaker[PDU], X)$ and CDU.DH contains a turn-initial "but" and CA_k: $disagree(speaker[CDU], CA_j)$ and CA_l: $assert(speaker[CDU], Y)$

   (a) If UDU.DH contains CA_i: $propose\_plan(Z)$ and TA interpreter returns [either $constraint(Z, X)$ OR $undesirable\_effect(Z, X)$] AND $alternatives(X, Y)$], then update CDU.DH with CA_m: $correct(speaker[CDU], CA_j)$

   (b) Else if UDU.DH contains CA_i: $assert(Z)$ where Z is a situation and UDU.TB[7] has $believes(speaker[CA_k], goal[\neg Z])$ and TA interpreter returns $\neg[effect(X, \neg Z)]$, then update CDU.DH with CA_m: $correct(speaker[CDU], CA_j)$

   (c) Else if calling TA interpreter with CA_j returns $undesirable\_effect(CA_j, CA_k)$ and TA interpreter called with Y and X returns $alternatives(X, Y)$, then update CDU.DH with CA_m:

[7] Task Belief field, discussed previously in Chapter 6.

$$correct(speaker[CDU], CA_j)$$

Notice that none of the above updates involve correction of discourse relations between commands and material in preceding turns; these sorts of discourse relations would be addressed in the section on correcting implicit information. Notice that we do not have fine-grained information within the *correct* CA as to what exactly is being corrected. However this is not as great a shortcoming as it seems, since we have a unique constellation of CAs (and TBs etc.) in each case, and distinguishing the various types of corrections of commands does not depend on just one CA. Also, we see what is disagreed with via the disagree CA and what is asserted as an alternative in the assert CA uttered by the corrector. In the next chapter, we will discuss how these distinct constellations of IS give sufficient information to generate an appropriate response depending on the hearer of the "but"s beliefs. Recall that here, as in the other cross-speaker "but" relations seen so far, a theorem-prover and planner function on the conversational and task related beliefs, expectations, plans, etc. of the speaker interpreting the turn. This enables different speakers to interpret dialogue differently depending on their own beliefs.

7.5 Denial and Rejection

Many of the examples in the previous section seemed to involve rejection of offers, negotiation (by introducing alternative possible actions), misunderstandings, argumentation (by proposing other arguments), etc. Correction involves proposing information that somehow invalidates (i.e., denies, rejects or disagrees with) the role of something in the preceding discourse, here restricted to the preceding turn, while asserting either why this previous information is invalid (in some command cases) or presents an alternative.

In the discussion on correction of assertions above we saw that what is corrected is what is actually asserted, while in the case of implicit information, what is corrected is some relation communicated either by the assertion (i.e., the discourse relation relating clauses within the assertion), or a relation between material in the current turn and the preceding discourse, or implicitly communicated expectations, etc.

This leads to a neat treatment of denial which falls out of the treatment of correction; we can take denial to be simply the rejection of the previous turn's assertion with no attempt at correcting (via new information) the denied information. We will also want to distinguish denial from rejection, which also rejects something in the preceding turn and fails to correct the rejected turn via new information. We can distinguish denial as countering the perceived truth-value of an assertion, while rejection turns down an offer, and is thus only seen in TOD, where commands are issued. So we also see that rejection rejects a command in the previous turn while denial rejects an assertion, and both fail to introduce new explanatory or corrective information.

For example, if C wants to deny B's answer in the embezzlement example, he simply needs to say something along the lines of "No he wasn't (caught embezzling funds from the pension plan)". A rejection of a command like "shut the door" would then simply involve an asserted "no!". We will not address denial or rejection further as they almost never involve "but", since they do not involve any acceptance beyond understanding of what is denied or rejected.

7.6 In Summary

Essentially we defined correction as involving disagreement with or unacceptability of something which is communicated in the previous turn, and the assertion of either an explanation or an alternative perceived to be more appropriate/relevant/correct by the corrector. We started out by presenting a range of corrections in assertions, answers to questions, questions and commands, presenting procedures describing how correction can be interpreted given differences in beliefs etc. discernible by the theorem-prover and planner. The theorem-prover and planner interpreted conflicts based on the corrector's beliefs, expectations, and plans w.r.t. the previous turn, enabling subjective interpretation. Despite the range of correction procedures presented, there are some trends which we can say typically characterise correction. In some cases we needed to use theorem-provers and planners to determine the conflicting beliefs in order to generate disagreement, and in other cases the disagreement was already present in the IS because the speaker had explicitly communicated it. In either case, the presence of

the disagreement/rejection/denial and, as mentioned earlier, either (1) an explanation of why the speaker disagrees or (2) an alternative proposal characterises correction.

A dialogue system modelling human-human conversations would not need to determine whether something is disagreed with or not, or if the hearer knows of conflicting/contradicting information and a more reasonable alternative, since the corrector will explicitly indicate this in order to correct the other speaker. In these cases, the system would only need to be able to detect disagreement and assertion CAs which provide either an alternative or an explanation and then check whether what is disagreed with and what is asserted pertain to the same topic in order to model what is communicated as correction.

In other words, given a constellation of interpreted CAs, correction is added to the constellation provided that the necessary criteria, i.e., disagreement with the preceding turn's claim and assertion of an alternative claim or explanation, are met. So the minimalistic claim, barring the more complicated use of a theorem-prover which can detect the presence of conflicting beliefs, would simply update the IS with correction given (1) disagreement with and (2) assertion of claims pertaining to the same subject, either in the role of anb explanation for the disagreement or presentation of an alternative to what was disagreed with, and could simply be modelled as:

1. If the current turn (in CDU.DH) contains a turn-initial "but" and CA_k: $assert(Z)$ AND ALSO [CA_l: $reject(speaker(CA_k),Y)$ OR $disagree(speaker(CA_k),Y)$], AND if the previous turn (PDU.DH) contains CA_j: $assert(Y)$ then pass the theorem-prover a pointer to the IS

   (a) If the theorem-prover returns $alternatives(Z,Y)$ OR $explanation(Z,Y)$, then update CDU.DH with CA_m: $correct(speaker[CA_k],CA_j)$

Notice that this approach simply looks for the necessary constellation of CAs in order to update the IS with correction, and summarises all correction cases seen thus far. One interesting and novel feature of correction as opposed to concession and Denial of Expectation is that there are no implicated expectations involved. In the next chapter we will summarise all three relations and distinguish how they differ.

Chapter 8

Conclusions

In this chapter we will draw the analysis together and summarise the relations considered again, presenting salient features and unique characteristics, and then put the analysis into practice by conducting an evaluation in which we examine some unseen examples from the corpora with the theory outlined thus far in mind. We finish by showing how DofE can be implemented in a working dialogue system.

8.1 Analysing Cross-Turn "But": a Summary

We will summarise the differences between the various relations addressed in the thesis, and then we will address where they overlap and situations in which relation assignment is ambiguous. We will then suggest how the various "but" signaled relations in dialogue can be distinguished via a single decision structure, in this case an update rule that distinguishes the various relations thus far addressed in both TOD and NTOD.

8.1.1 Reviewing Characteristics of "But" Signaled Relations

We will start by reviewing the various relations addressed in this thesis thus far and highlight their differences. Considering Denial of Expectation (DofE), concession and correction, we will start by summarising their characteristics again; assume B represents the speaker of the "but" turn and A the preceding speaker below:

- *DofE*: B infers a defeasible rule ($p > q$) underlying the preceding turn's assertion (A) in which A's assertion forms the left-hand-side (LHS, p) of the rule and B believes that the RHS (q) does not hold, contrary to expectation, and asserts this. B either accepts A's assertion or understands it and disagrees, giving evidence for disagreement by denying q. DofE in TOD is seen as involving a frustrated plan rather than a denied expectation (as in NTOD).

- *Concession*: Both speakers make points related to a given salient topic (i.e., the TC), and B infers that A's turn relates to this topic and makes a different point w.r.t. this topic. Here B either agrees with or (minimally) acknowledges (possibly implicitly) A's utterance.

- *Correction*: The corrector (B) rejects/denies all or part of A's turn and then makes an alternative relevant claim on the same topic to either (1) replace what A communicated or (2) explain why PDU is rejected.

So now that we have summarised how we term these three relations, we can see how they are different:

- Concession involves acceptance or minimally acknowledgement of the previous turn while correction involves rejection of or disagreement with the previous turn; DofE can involve either acceptance or rejection.

- While DofE and concession necessarily involve defeasible rules, correction simply involves the corrector's perspective on something wrong about the preceding turn and the offer of an alternative or an explanation, with no implicatures of any consequent.

- DofE involves inferring an expectation of a consequent that does not necessarily have a basis in the dialogue history, unlike concession, which involves inferring the stance of the preceding turn towards a conversationally established topic (i.e., the TC); the consequent of the DofE rule arises solely due to the beliefs of the hearer of the DofE, and is therefore more debatable.

- Concession (and to an extent correction) involves stance and perspective w.r.t. a salient claim much more than DofE.

- Both correction and concession involve introduction of new and relevant information; the concessive turn puts forward a new perspective w.r.t. the TC which does not deny the truth of the preceding turn, but merely presents a (possibly favored by the speaker) alternative, while the correction is based on the speaker's denial/rejection of information in the previous turn and introduces a replacement for information in the previous turn.

We will briefly touch upon some ambiguous cases we have seen so far before reviewing the main criteria for distinguishing the three relations we focus on in this thesis.

One issue involves determining how we can distinguish between concessive and correction cases featuring a question followed by two assertions. If we see a *what* question for example, it can easily be followed by a pair of alternative perspectives, but they will often also propose alternative answers, e.g.,

(8.1)
B: What would you like to do on this trip?
A: Well I'd like to take a long train trip through the mountains.
B: But that wouldn't leave us time to go swimming at the beach.

The difference between concession and correction hinges on the fact that concession involves acceptance or minimally acknowledgement of the previous turn while correction involves rejection of or disagreement with something in the previous turn. Furthermore, concession involves conversational implicature relating the propositions in the turns with a relevant claim while correction does not involve implicature, but rather denies something in the discourse history while asserting a replacement or explanation. Seemingly ambiguous, corrections of answers appear very similar to a pair of concessively related assertion turns following a question, since the question being answered would serve as the central proposition or claim (i.e., TC) involved in concession. However in corrections, the correction provides a replacement for the prior answer, unlike concession, which (in the case of a preceding question) simply provides alternatives.

Also, concession following a question does not involve the denial of the previous turn, and putting forward an alternative perspective is not the same thing as correcting something. In the example above, the question does not ask for a factual answer but rather one based on the speaker's inclinations, which makes correction based on an-

swerhood (i.e., the correctness of the answer) somewhat less clear. If B corrects A in this example, she can only really correct factual information relating to A's assertion, e.g., by introducing a constraint, precondition failure, denying the truth of something communicated in PDU, or defeating a presupposition. We model this case as involving the acceptance of (at least part of) the previous turn and the assertion of new information w.r.t. the question that the speaker possibly deems more relevant, which is concession. This example illustrates that in order to correct attitudinal answers the corrector is most likely to find fault with factual information or presuppositions contained in the correctee's answer rather than its answerhood; one cannot correct another's opinion, only presuppositions contained in that opinion can be corrected.

The difference between interpreting concessive and corrective responses depends on whether something in PDU is denied or not, and then, whether this denied element is argued for in a different way in PDU (i.e., whether it is a TC). A particularly nebulous area between these two relations is when speakers argues for and against a TC. We will assume that these cases, like Example 6.6 in Chapter 6 (i.e., the TRAINS example, reprinted below)

(8.2) U: well actually I'll already have an engine in Bath after I unload the boxcars right p
S: right but you wouldn't have it in time q
Concession: TC=plan is OK; $X = precondition\_met(p, subgoal(TC))$;
$Y = precondition\_failure(q, subgoal(TC))$

are concessive. This is because concessive cases like these do not involve rejection of anything explicitly present in PDU, but rather deny that a shared goal in the speaker's beliefs will hold given the inference that the speakers' goal will be met. This is a tricky case because the inference is denied while the previous turn's assertion is accepted, which makes this seem like a candidate for correction of implicit information. However notice that the inference involves the TC (i.e., the success of their shared goal) which is the topic of both turns. Recall that concession does not deny the previous turn itself, and if there is a TC which is being argued for and against, this is concession despite the presence of an argument against, which can also be seen as denying the truth or validity of the TC. Consider Example 8.3 (reprinted from Chapter 7), another potentially difficult case:

(8.3) A: Chairs have four legs for stability.
B: But three-legged chairs are equally stable, they just need to be well-designed.

We claimed that B corrected PDU (A's turn) here since B asserts new information on the same topic (stability of chairs) while correcting A's assumption that the reason chairs have four legs is for stability. More specifically, what is corrected is the explanation being communicated by A, not the assertion itself. B probably agrees with A's assertion, just not the explanation that chair-stability requires four-legs. Notice that what is being corrected is in the previous turn itself, not in some expectation that holds given the previous turn's assertion and therefore this is not DofE. Given a suitable TC in the dialogue history however, a concessive reading is possible. For example, if the TC were "stability depends on having four legs", we could see A as exemplifying this TC by asserting that chairs have four legs for stability, and B countering this TC via a counter-example (i.e., three-legged chairs). So we see that cases like these are distinguished by the presence of a TC; if there is a TC which both turns relate to and CDU does not explicitly deny or reject the aspect of PDU it provides an alternative perspective on (w.r.t. the shared TC), then this is concession.

Correction can also involve correcting a relation between the previous turn and something else further back in the dialogue history, like answerhood, as we discussed earlier, but it does not involve simply denying an inference that arises from the previous turn's utterance itself; if a correction were to correct such an inference, it would have to deny the inference and then either give an explanation or an alternative. However it does not make much sense to think of alternative inferences that can arise from PDU. So the grey area between DofE and correction occurs in cases which involve introduction of an alternative or an explanation. These cases can either be characterised as (1) DofE with either (a) explanation of DofE or (b) presentation of an alternative inference (RHS) given PDU, or they can be seen as (2) the correction of an inference arising from PDU and thus involving denial and either an explanation or an alternative. Simple denials of inferences from PDU are DofE.

Concession will likewise be distinguished from DofE via the distinction made in Grote *et al.* (1995), since there will be a TC argued for and against in the turns. Additionally, in concession the "but" turn will not be denying an expectation that arises solely via the previous turn in his beliefs/plan, but rather will argue either for or against

a conversationally established claim. This means that a TC should be grounded and both turns should relate to this claim rather than directly to each other in concession.

8.1.2 Summarising Distinguishing Features of the Analysis

We will review briefly below the main criteria for interpreting the three relations addressed in this thesis before turning in the next section to an evaluation of the analysis on unseen data (A refers to PDU and B to CDU below):

- *DofE:* interpreting DofE requires finding an inference rule in either the hearer's (A's) Private Beliefs (in NTOD) or in their task-plan (TOD) where the LHS of the rule is A's assertion, and the RHS comes from the negation of CDU (B's turn). B either implicitly accepts A or understands and disagrees where $\neg q$ in inferred rule $p > q$ is evidence of why $\neg A$. The inferred rule arises from B's mental state and not from the DH, unlike concession. TOD involves a frustrated plan/task-related belief rather than a denied expectation as in NTOD.

- *Concession:* interpreting concession involves determining relations X and Y between the propositions contained in the two turns sandwiching the "but" and the TC (i.e., a contextually relevant claim). We postulate concession for any constellations of X and Y relation pairs provided that B agrees or minimally understands and acknowledges A (and X) while contributing novel information w.r.t. the TC. Additionally some cases involve a further relation Z communicated by B which holds between the two relations X and Y.

- *Correction:*

  - Correcting Assertions: conflicting/contradicting information leads to disagreement with and rejection of this information and correction via assertion of new and relevant information (either as an alternative or an explanation to what is rejected); as in the other cases, presuppositions contained in PDU can also be corrected by signalling rejection and giving an explanation or providing an alternative

- Correcting Answers to Questions: both turns provide alternate answers and the second turn rejects the answerhood of the first; the assertion in the first turn is either explicitly or implicitly accepted
- Correcting Implicit Information (e.g., DofE, concession, alternatives, propose plan): the relation is rejected and an alternative is asserted
- Correcting Questions: either the question can be found to be irrelevant or something in the question is rejected (either by introducing a precondition/constraint that must be overcome or by giving an alternative) and an assertion on the topic (either an alternative or an explanation) is communicated
- Correcting Commands: here the command is rejected and a constraint or undesirable effect of the command is introduced as an explanation

And now we will consider how interpreting these relations affects deliberation of a response from the hearer:

- *DofE:* check hearer's beliefs/task-plan to see if the rule and assertion exist there and respond accordingly
- *Concession:* hearer can dis/agree partially/completely with explicit/implicit content or relevance of *concession* or *alternatives* relation, *criticise* or *implicit_accept* CAs, task-planning operators communicated, *alt_opinions* or *alt_beliefs* or *invalid_action* updates. Disagreement can be expressed via presenting alternatives or via criticism
- *Correction:* hearer can dis/agree partially/completely with explicit/implicit content or relevance of what is asserted, corrected, and/or disagreed with in the preceding turn

Now that we have firmly established what makes these three "but" triggered cross-turn relations unique, we will discuss how the three "but" relations' procedures fit into an overall procedure for addressing turn-initial "but". The distinctions between the relations can best be captured by the decision tree in Figure 8.1 which illustrates

the choices one must make in distinguishing relations. Notice that acceptance (or rejection) of what is explicitly stated in PDU is the first choice factor. Disagreement with what is explicitly conveyed (i.e., not with inferences, relations, etc. arising from PDU) and the presence in CDU of an alternative or explanation is clearly correction. If something other than an alternative or explanation is asserted, and the expectation $PDU > \neg CDU$ is assumed to hold in B's (the speaker of the "but"s) beliefs, then this is DofE. If on the other hand, CDU expresses agreement or acknowledgement of what is explicitly stated in PDU, one must distinguish between acceptance or acknowledgement and rejection of implicitly conveyed material like relations (e.g., answerhood), inferences, presuppositions, etc. If the implicitly conveyed material in PDU is accepted or acknowledged, we have concession. If on the other hand the implicitly conveyed material is disagreed with, the choice depends on whether there is a TC which both turns are related to. In the case of a TC which both turns are related to, we have concession. If on the other hand CDU presents an alternative or explanation for the implicit information conveyed in PDU that it rejects, then we have correction.

The decision tree in Figure 8.1 represents the choices a procedure modelling turn-initial "but" should make to distinguish the relations. This decision structure should take the form of an update-rule triggered by turn-initial "but" which accounts for these distinctions. It should be noted that this decision tree summarises the distinctions made in the relations' procedures very roughly, and is not a replacement for the specific distinctions and subsequent updates outlined in the relations' procedures proposed in the thesis. The idea here is to simply sketch how the space of "but"-conveyed relations is roughly partitioned. In the next section when we present an evaluation of examples, we will present this decision structure as a procedure which we will use to guide our distinctions in the evaluation.For the update procedure using this decision structure, we intend that if (for example) DofE is predicted by the decision structure, then the DofE procedure outlined in Chapter 5 is called to verify the details of the DofE relation and update the IS appropriately. That is, the decision should not be made on the basis of the decision structure alone, but in conjunction with the procedures outlined in the thesis. The decision structure simply prevents the interpretation procedure from inefficiently testing relations which cannot hold given the particular dialogue's characteristics. So

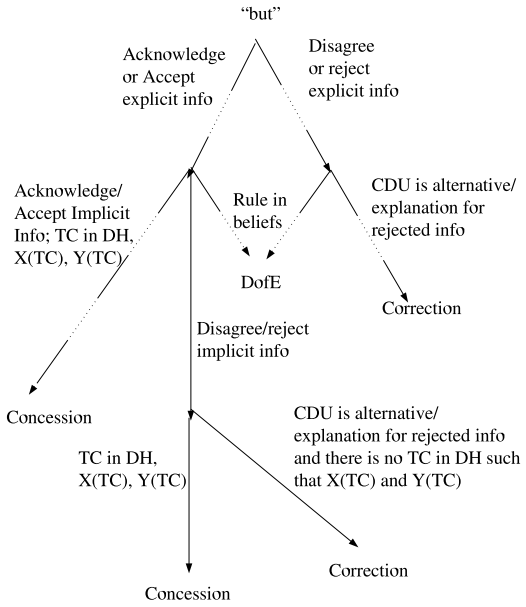

Figure 8.1: Decision Tree for Interpreting "but" Relations

for example if a dialogue with turn-initial "but" in CDU comes back from the CA interpreter with acceptance of both explicitly and implicitly communicated information in PDU, DofE and concession procedures should be called and tested. We want to emphasise that the decision structure is intended to roughly summarise the distinctions made in the proposed theory outlined in the thesis and account for this theory's characterisations of "but"-signalled cross-turn relations.

Since multiple interpretations may hold, branches in the decision tree do not explicitly exclude continuing down the tree. For example, if explicitly conveyed material is accepted and the interpreter can reasonably assume that the speaker of the "but" has a belief of the form $PDU > \neg CDU$, then this is DofE. But CDU might also accept implicitly communicated information, in which case, this example can also be labelled as concession. In other words, the decision tree should be explored exhaustively from the first branch which succeeds on down that subtree.

After the relations' interpretation procedures have been called, they result in the update of the IS with the appropriate information. We assume that these IS updates are reasonable if the analysis is correct, since the updates basically note that the relation holds, along with additional inferences involved and any resulting socially committed propositions (SCPs in the IS). Then, after the other update rules have been called, the last update stage involves deliberating on what the hearer of the "but" can say. Depending on the relations the IS was updated with from this "but", the respective relations' deliberation procedures should be called.

8.2 Evaluation

Bearing the summary of the relations' characteristics in mind, we now turn to look at twenty-one unseen examples from the TRAINS, Switchboard, Maptask, Autoroute and AMEX corpora (the latter two are both travel reservation domains). These examples were chosen because they appeared to contain cross-turn "but" and are the first twenty-one cases of cross-turn "but" found.

The goal of this section is to evaluate the proposed methodology by (1) validating the functionality (i.e., testing whether the methodology works as intended on new ex-

amples) and (2) determining if it makes good predictions about unseen data. In the absence of automatic means of evaluating our analysis we evaluate these examples manually, running them through the decision tree in Figure 8.1 in order to determine whether the proposed methodology makes sensible predictions on this unseen data. As stated earlier, the decision tree represents the distinctions between relations made in the theory and thus acts as a tool by which we can distinguish relations in the evaluation and also in the update rule. That is, decisions made via the decision tree reflect the thesis' treatment of these "but" cued relations and enable us to evaluate the thesis on these unseen examples.

8.2.1 Analysing the Data

We analyse these examples following a simplified scheme; we begin by analysing the examples, (a) noting salient propositions and SAs communicated in PDU and CDU, then (b) noting what is accepted/acknowledged/rejected in PDU and noting if this is implicitly done, and finally (c) noting alternatives presented and what they are alternative to, or noting whatever is communicated otherwise (e.g., explanations). These examples are then classified according to the decision structure in Figure 8.1 which is outlined in the procedure below.

1. If B (in CDU) rejects material explicitly stated in PDU (A's turn) then
    (a) If CDU also provides an alternative replacement or explanation for what is rejected, this is correction
    (b) If CDU expresses something that is not an alternative or explanation for PDU and the hearer interprets that the rule $PDU > \neg CDU$ holds in B's beliefs, this is DofE
2. Else if CDU accepts/acknowledges material explicitly stated in PDU then
    (a) If CDU accepts/acknowledges material implicitly communicated in PDU, this is concession
    (b) Else if CDU disagrees with/rejects material implicitly communicated in PDU, then

i. If there is a TC in the DH such that relations $X(CDU, TC)$ and $Y(PDU, TC)$ hold, then this is concession

ii. Else if there is no TC which both turns are related to and CDU presents an alternative or explanation for the implicit information in PDU which it rejects, this is correction

The evaluation below has not been validated by other coders. Nonetheless it is our hope that the simplicity of the decision structure and especially its reliance on relatively unambiguous characteristics of the dialogue should speak for its reproducibility. Specifically, noting salient propositions, SAs, what (in PDU) is accepted or rejected by CDU, and whether alternatives are presented are relatively straightforward. Implicit acceptance and rejection may be somewhat more debatable, but this is also the case for humans engaging in dialogue; one must infer acceptance/rejection in cases without explicit signals and then interpret on the basis of this inference. These interpretations are defeasible, and in dialogue the hearer can refute wrongly attributed relations if she infers that the speaker has inferred the relation from her turn. Also, despite the fact that we do not refer to a specific annotation scheme and there are many possible SAs one can assign to an utterance, this is not a significant problem since the distinction procedure does not rely on SAs to distinguish relations. Turning to the decision procedure itself, we will address in the subsection following the evaluation isssues involving how the TC is determined, given that it may not be explicitly established in the DH, and likewise the difficulties of assuming that an expectation reasonably holds in the speaker's beliefs for DofE.

We present the evaluation below, ordering the examples by their classification so readers can see commonalities between cases that are classified alike.

8.2.1.1 Concession Examples:

- Switchboard index 14203#MATCH[1]:

(8.4)
B: [[Texas is known for it's poor schools] and [[that's too bad] because [we got little kids]] and [we're not too pleased about that]]
A: But we're not funding them like other states either though

[1]All Switchboard dialogues can be found at http://www.ldc.upenn.edu/cgi-bin/lol/swb/. The index provided with the examples is the dialogue identifier.

Analysis:

1. PDU: assert[Texas is known for it's poor schools];
   CDU: assert[we're not funding them like other states]
2. CDU accepts PDU's assertion and doesn't directly deny/disagree with PDU
3. alternative perspectives on why [Texas has problems]
4. Classification: concession: X=exemplification(PDU,[Texas has problems]), Y=explanation(CDU,[Texas has problems])

- Switchboard index19940#MATCH:

(8.5) A: oh [so that's how you knew] [[well we're first time homeowners] and [I'm still scared about everything like that going wrong]] and [how do you know it's going to happen and all]
B: yeah but [[I think so far as the locations richardson plano are probably comparable] [the school districts are about the same quality wise]]

Analysis:

1. PDU: assert[well we're first time homeowners and I'm still scared about everything like that going wrong];
   CDU: assert[I think so far as the locations richardson plano are probably comparable]
2. CDU explicitly accepts PDU; nothing is explicitly denied/disagreed with
3. alternative perspective with respect to buying the new house is put forward in CDU; here PDU explains why he is worried about buying the new house while PDU asserts info which mitigates PDU's worry
4. Classification: concession: TC=[buying the new house is scary];
   X=explains(PDU,TC); Y=counter-evidence(CDU,TC)

This is clearly concession, with no disagreement and simply involves putting forward alternative perspectives w.r.t. the TC that buying a new house is scary.

- Transcription for American Express[2] (AMEX) tape 1: A 4-24-89:

  (8.6) A: if that's the best we can do. what, just for curiosity if you stayed over Saturday what are the rates?
  B: well we could get it down to a hundred and twenty four dollars round trip
  A (PDU): round trip ok
  B (CDU): but that w- but that would be completely non refundable

  Analysis:

  1. PDU: assert[the round trip is ok]=accept[plan so far];
     CDU: assert[that would be completely nonrefundable]=introduce- condition of plan
  2. CDU implicitly acknowledges PDU's acceptance of trip and nothing is denied/
     disagreed with
  3. w.r.t. claim that the flight offer is good, PDU accepts claim while CDU qualifies
  4. Classification: concession: TC=[flight offer is good]; X=accept(PDU,TC); Y=introduce-condition/qualification(CDU,TC)

  Notice that this example almost does not count as a cross-turn "but" if one takes PDU's turn as a back-channel (indicating only understanding). If this were the case, this example becomes an intra-turn (same speaker) medial "but" utterance. However if we instead see this as involving A indicating acceptance of the previous turn's flight offer, then the analysis as concession works fine, as there is no disagreement, and accepting the offer is in contrast with B's introduction of a qualification/constraint to the offer in CDU.

- AMEX Tape 7:

  (8.7) A: Umm ok San Francisco Tokyo uhh fare quote for business round trip would be twenty one fourteen. and again the coach was seventeen fourteen. and then for the

[2] See http://www.ai.sri.com/~ communic/amex/amex1.html.

twenty ninety nine that was economy but there it is also offered for business. and first class the business around the world fare would be twenty nine ninety nine and first class thirty nine ninety nine. and that's, again the same of of applications of four stop overs uh for minimum maximum of fifteen. and the purchase restrictions were twenty one days in advance cancellation within ah or longer then twenty one days prior to the departure you get a full refund, within twenty one days of the flight it's twenty five percent penalty, umm attached to all all of those fares, economy, business and first and if umm bu- in case of illness or death or anything like that it is uh waivable, the twenty five percent [sick clause on requirements]
B: But that even) applies to these apex and stuff

Analysis:

1. PDU: assert[the purchase restrictions were....sick clause on requirements]; CDU: question[that applies to these apex and stuff] (assuming rising intonation)
2. CDU acknowledges PDU and nothing is rejected/disagreed with
3. w.r.t. claim that flight conditions are specified, CDU raises a question
4. Classification: concession: TC=[flight conditions are specified]; X=explain/elaborate(CDU,TC); Y=question-applicability(CDU,TC)

SA plays a stronger role here than in many other cases. This is because the question introduces doubt and raises a perspective with a contrasting polarity to X.

- AMEX Tape 7:

(8.8) A: (PDU) there are no penalties for cancellation if changes are made it's subject to whatever the new fares are at that time
B: ok
A: (PDU) cause they're going to increase the fare
B: (CDU) but there's no penalty existing

Analysis:

1. PDU: assert[there are no penalties for cancellation if changes are made it's subject to whatever the new fares are at that time]; assert[they are going to increase the fare];
   CDU: question[there's no penalty existing] (assuming a rising tone)
2. CDU implicitly acknowledges PDU; nothing is rejected/disagreed with
3. CDU raises an alternative clarification w.r.t. cancellation penalties
4. Classification: concession: TC=[the cancellation penalties are N]; X=elaborate/explain/answer(PDU,TC); Y=clarification-request(CDU,TC)

This is another example in which concession captures the alternative perspectives w.r.t. the salient claim (or TC) which is under discussion.

- AMEX Tape 13:

(8.9) A (PDU): I I would need that code number to apply for that conference rate
B (CDU): ok but then that's two hundred dollars more than this non refundable

Analysis:

1. PDU: assert[I need that code number to apply for that conference rate];
   CDU: assert[then that's 200 dollars more than this non refundable]
2. CDU accepts PDU and doesn't disagree/deny anything
3. CDU notes a downside to conference rate as compared to the nonrefundable ticket
4. Classification: concession: TC=[we will apply for conference rate]; X=precondition(PDU,TC); Y=negative-comparison(CDU,TC); Z=prefer-alternative(CDU,PDU)

- AMEX Tape 13:

(8.10) A (PDU): uh whereas th- this non refundable is non refundable but if it's you know it is two hundred dollars more if he can a go with the penalty he's going to come out ahead with the um the non conference rate
B (CDU): but I could I could still give you this Delta fi le code number and and I could

do it through you instead of having to deal with these people down in Saint Petersburg Beach

Analysis:

1. PDU: assert[this is non refundable but if it is 200 dollars more if he can go with the penalty he's going to come out ahead with the non conference rate];
   CDU: assert[I could still give you this Delta file code number and I could do it through you instead of having to deal with these people down in St. P Beach]
2. CDU implicitly acknowledges PDU; and doesn't disagree/deny anything
3. CDU presents an alternative course of action to the one PDU proposes
4. Classification: concession: TC=[we should take the nonconference rate option]; X=supporting-evidence(PDU,TC); Y=propose-alternative(CDU,TC)

Notice how the two turns are clearly related to a third claim (i.e., the TC), rather than directly to each other. This debate surrounding a central claim is what characterises cross-speaker concession.

- AMEX Tape 15:

(8.11) A (PDU): you know by the end of s- like August some markets start to, you could probably at least still get it in September you know probably by October though you'll fi nd it really diffi cult to fi nd the excursion space
B (CDU): But (I should defi nitely
A (interrupts): the most the most) inexpensive ah
B (CDU cont'd): I should defi nitely do it in July?

Analysis:

1. PDU: assert[you could probably at least still get it in September ...excursion space]
   CDU: question[I should definitely do it in July]

2. CDU implicitly accepts PDU's assertion and doesn't disagree/reject anything
3. CDU questions a necessary precondition
4. Classification: concession: TC=[client should take the excursion space in the summer]; X=introduce-constraint[PDU,TC]; Y=request-confirmation-precondition[CDU,TC]

PDU introduces the constraint that the excursion space is only valid until and including September, while CDU accepts this and requests confirmation of the precondition by asking if July is the time to take the trip.

- Switchboard index 20073#MATCH:

(8.12) A (PDU): in our neighborhood because all of the homes did have these foundation problems and we didn't know um you know if we were afraid we'd be living there and all of a sudden the house would crack in half and split open
B (CDU): um-hum
B (CDU cont'd): yeah BUT uh well our house and i'm not sure how all of the b[uilders]- you know how most of the builders do but ours is what they call uh post tensioned which means that there are iron bars you know that run through the foundation

Analysis:

1. PDU: assert[in our neighborhood all the homes have these foundation problems and we were afraid the house would crack in half]
   CDU: assert[our house is post tensioned which means there are iron bars that run through the foundation]
2. CDU explicitly accepts PDU and doesn't disagree/reject anything
3. CDU presents alternative perspective w.r.t. foundation strength
4. Classification: concession: TC=[foundations (in general) are strong and stable]; X=counter-evidence(PDU,TC); Y=supporting-evidence(CDU,TC)

- Switchboard index 22185#MATCH:

(8.13) B: i know when my mother was a you know going into the work force there wasn't very many opportunities for her i guess she's in her late forties you were expected to stay home and take care of the kids and i've never faced that at all
A: um no i think now now you go to work when they're six weeks old you know and you stay there and i feel like the next generation you'll probably just work right along with your husband you know
B (PDU): plus i bet it cuts cuts down on your absenteeism because you've got two days off that you can do everything so you don't have to you know unless you're just really sick or the child's sick
A (CDU): but yeah yeah because most time now you know it's just a weekend or just forget it yeah-

Analysis:

1. PDU: assert[it cuts down on your absenteeism because you've got 2 days off that you can do everything so you don't have to unless you're sick or the child's sick]
   CDU: assert[most time now it's just a weekend or forget it]
2. CDU explicitly accepts PDU
3. CDU provides alternative evidence on the utility of the weekend off
4. Classification: concession: TC=[working mothers fare better now]; X=supporting-argument(PDU,TC); Y=express-utility(CDU,TC)

- Switchboard index 30343#MATCH:

(8.14) A (PDU): that's got us in a little bit of trouble now that's why we and we was using one of them to pay off the bills on the others and that got that in trouble there
B (CDU): yeah yeah
B (CDU): but uh we uh we try to keep a uh
B (CDU): uh tight controls over over them but it it gets hard like especially around christmas time and birthdays

Analysis:

1. PDU: assert[that got us in trouble so that's why we were using one of them to pay off the bills on the others]

CDU: assert[we try to keep tight controls over them but it gets hard around Xmas and b'days]

2. CDU explicitly agrees with PDU and doesn't explicitly disagree
3. CDU appears to put forward an alternative perspective
4. Classification: concession: TC=[Situation causes trouble]; X=explain-solution (PDU,TC); Y=supporting-evidence(CDU,TC)

Notice how difficult it is to determine a specific TC in some cases without more dialogue context; this is not unexpected, as DH gives rise to the TC and this was omitted in this example. The Situation mentioned in the TC refers to the problematic situation the speakers are talking about.

- Maptask Dialogue[3] q1nc1:

(8.15) GIVER[1]: mm , and ha– have you to go have you to go there like in ... you've not you've not to pass the ca– you've not to pass that fence ?
FOLLOWER[2]: well , if i went three inches down from the caravan park ... ... i would ... you know i'd more or less hit the picket fence .
GIVER[3]: so then , right i see , so is that the idea of this then so you you're going straight to where i am ... instead of going round the picket fence ?
FOLLOWER[4]: where's your ?
GIVER[5]: if i say if i say that you've to go to the old mill ... fi rst .
FOLLOWER[6]: actually go to it ?
GIVER[7]: mmhmm .
FOLLOWER (PDU): well, i ... you know i wouldn't have to really go down i could just i could just go ... you know ... down and to the right rather than go down and then turn right sort of thing.
GIVER (CDU): hmm, but is the idea is the idea ... so that you don't go there to the picket fence ... if i'm going to the old mill fi rst?

Analysis:

1. PDU: assert[i could go down and to the right (diagonally) rather than go down and turn right]

[3]http://www.ltg.ed.ac.uk/ amyi/maptask/demo.html

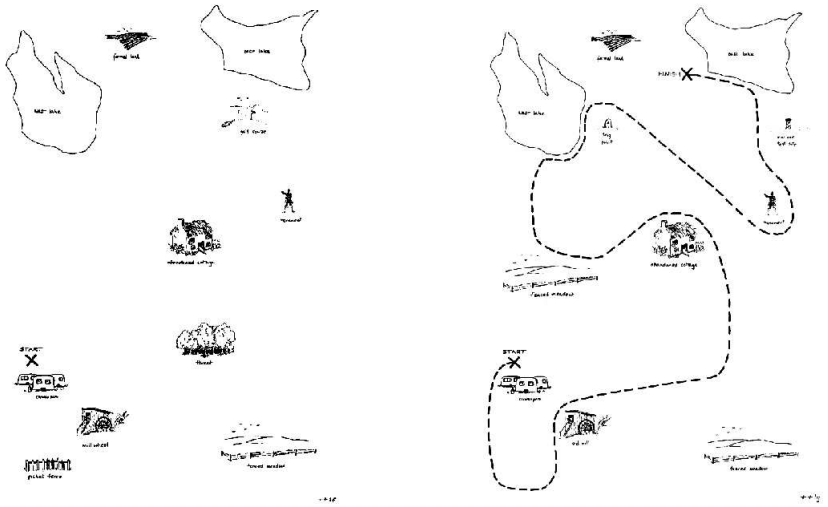

Figure 8.2: Follower map (left) and Giver (right) map for Example 8.15

CDU: question[is the idea that you don't go to the picket fence if i'm going to the mill first]

2. CDU implicitly accepts or at least acknowledges CDU and doesn't disagree
3. CDU requests clarification on the goal behind PDU's assertion
4. Classification: concession: TC=[goal is to go down and to the right]; X=suggest-manner-action[PDU, TC]; Y=request-clarification[CDU,TC]

The Giver does not have the picket fence referred to in FOLLOWER[2] and his map includes a path that goes down from Start and across and up to the old mill. GIVER[3] tries to figure out whether Follower should just come directly to him rather than go down and across and up to the old mill like the path on his map shows. PDU proposes going diagonally to the old mill rather than down and across, since she learned from GIVER[5] that she must get to the old mill and the picket fence is in the way on her map if she goes down and across and up as Giver suggests. When she mentions the picket fence in [2], the Follower picks up on this and clarifies that she wants to avoid the picket fence down from where

she is in [3]. However the fact that she is trying to get to the old mill in the end does not affect the TC radically in this case since what is being debated here is how the Follower should get to the old mill, and her turn (PDU) proposes going diagonally, which the Giver tries to clarify, since he does not have the picket fence on his map and probably wonders why she proposes this alternative way of getting to the old mill.

- Autoroute 2/5 Woz Tape 5: side 1; Dialogue 9:

(8.16) W: Please wait while your route from Hereford to Darlington is calculated.
W: The quickest route is two hundred and forty miles and will take four hours nine minutes.
W (PDU): Would you like me to send the instructions to you.
C (CDU): Uhm, yeah, but what I'd- (.) another thing I'm interested in is- is going via Sheffi eld, how near dis- to Sheffi eld does that uh (.) route take me.

Analysis:

1. PDU: question[would you like me to send instructions to you]
   CDU: answer[yes]; assert[what i'm interested in is going via sheffield]; question[how near sheffield does that route take me]
2. CDU accepts PDU's offer and asserts that she is interested in an alternative route
3. Classification: concession: TC=[route instructions are sorted];
   X=offer(PDU,TC); Y=request-clarification[CDU,TC]

Another possible way of analysing this example would be to see *X* as implicitly communicating that the TC holds, because the instructions must be sorted out before they can be sent (this is a precondition), and Y implicitly communicates that the TC does not hold, because he does not know yet whether the route goes via Sheffield, and if it does not, he wants a new route that does go via Sheffield.

- Autoroute 2/5 Woz Tape 5: side 1; Dialogue 11:

(8.17) W Please wait while I set that preference and recalculate your route.
W The best route avoiding motorways will take three hours and fi fty minutes.

C (PDU): Will that route be okay for a caravan.
W (CDU): Yes, but it will take longer because the speed restrictions on a caravan are different.

Analysis:

1. PDU: question[will that route be ok for a caravan]
   CDU: answer[yes]; assert[it will take longer because the speed restrictions on a caravan are different]
2. CDU affirmatively answers PDU's question and provides a downside
3. Classification: concession: TC=[the route will be ok for a caravan]; X=question[PDU,TC]; Y=explain-negative-effect[CDU,TC]

8.2.1.2 Correction Examples:

- Maptask Dialogue q1nc2:

(8.18) FOLLOWER: a vertical line up ?
GIVER: yeah , maybe about another two and a half inches .
FOLLOWER: so i should be ... f– more than halfway up the page further ?
GIVER: are you sh– , you should be about exactly halfway up the page , and but you're still kind of at the right-hand side .
FOLLOWER: i know, i'm at the i'm about two and a half inches off the right-hand side.
GIVER (PDU): right that's fi ne.
FOLLOWER (CDU): but i'm ... mm but i'm too high i think ... ... right doesn't matter just.

Analysis:

1. PDU: accept[FOLLOWER[i'm about 2.5 inches off the right-hand side]]
   CDU: assert[i'm too high]
2. CDU disagrees with PDU's assessment that he is at a good place and says why

3. Classification: correction.

Note that the correction algorithm would need to interpret PDU as assuming that this (2.5") is acceptable, which could be found by searching his beliefs or planning given his beliefs. It is this assumption which is being corrected by CDU.

- Maptask Dialogue q1nc1:

(8.19) FOLLOWER: but the mill wheel's to the right of it .
GIVER: uh-huh ... that's fi ne , so if you you're going to the old mill .
FOLLOWER: fi rst of all ?
GIVER: mmhmm .
FOLLOWER: it's the very fi rst place i go ?
GIVER: uh-huh .
FOLLOWER (PDU): well, i can just draw a line straight between the two then, right?
GIVER (CDU): yeah but on my map it's not got that it's eh, you're going in a... ... sort of curve ... to the bottom of the map then round the old mill.

Analysis:

1. PDU: question[i can draw a line straight between the two]
   CDU: assert[on my map it's not got that...you're going in a curve to the bottom of the map then round the old mill]
2. CDU explicitly answers PDU's question affirmatively but disagrees with PDU's characterisation of the problem
3. CDU adds clarification to this answer
4. Classification: correction: states what is wrong (that a straight line can be drawn between the two) and then explains what the map really looks like.

In this case we rule out the concessive interpretation because concession should not reject anything and should minimally provide implicit acceptance. So we simply model this as correction, which gets at both what is wrong (i.e., that a straight line can be drawn between the two) and explains why (i.e., you're going in a curve to the bottom of the map and then round the old mill).

- TRAINS d93-12.4:

(8.20)
utt97 u: okay um and to make the orange juice orange juice and load the tankers
utt98 s: yep
utt99 u: uh that's an hour
utt100 s: um yep so that'll be six p.m.
utt101 u: wait a minute where am i now oh we're done
utt102 s: oh but I thought we had to get the orange juice to Avon

Analysis:

1. PDU: assert[task is done];
   CDU: assert[we have to get the OJ to Avon]
2. CDU implicitly rejects PDU's assertion by introducing an action which still needs to be done
3. CDU notes alternative to PDU's proposition by introducing a new action to be done
4. Classification: correction

Recall that anything which involves rejection and provides an alternative or explanation is correction. In this example CDU directly refutes A's claim that the task is done and gives an explanation (i.e., that something more must be done), which is clearly correction.

8.2.1.3 Examples which are both DofE and Concession:

- AMEX Tape 21:

(8.21)
A (PDU): ok well right now I'm working within ah within your (reservation
B (request clarifi cation): within my reservation
A (clarify): yeah
B (CDU): but but ah I'll have E. call because he may want to make some adjustments

Analysis:

1. PDU: assert[right now I'm working within your reservation]
   CDU: assert[I'll have E call because he may want to make some adjustments]
2. CDU implicitly accepts PDU but introduces new action/precondition presumably because she infers that PDU > reservation will be finalised soon, and all necessary information is present to accomplish this; presumably this is disagreed with, hence the introduction of a new precondition
3. Classification: DofE: [right now I'm working within your reservation] > [necessary information to finalise reservation is present];
   concession: TC=[reservation is nearly finalised]; X=provide-evidence(PDU,TC);
   Y=introduce-new-precondition(CDU,TC)

This example cannot be seen as exclusively concessive, because we can see CDU as directly countering an inference that arises from PDU, and assuming this expectation is in PDU's task-beliefs, he can signal DofE if he knows that the inference does not hold.

- Switchboard index 22185#MATCH:

(8.22) B (PDU): uh-huh i guess i take it for granted kind of because i just it's always been that way but i know
A (agree): um-hum
A (CDU): yeah but i can remem[ber]- remember back growing up my mother i mean it was she always worked but it wasn't that easy for her to just take off

Analysis:

1. PDU: assert[i take it for granted because it's always been that way]
   CDU: assert[my mother always worked but it wasn't that easy for her to take off]
2. CDU explicitly agrees with PDU but adds new information
3. Classification: DofE: [always experiencing good leave policies at work] > [there's always been good leave policies at work]

concession: TC=[there's always been good leave policies at work]
X=supporting-evidence(PDU,TC); Y=counter-evidence(CDU,TC)

8.2.1.4 Unclassified Examples:

- Maptask Dialogue q1nc2:

(8.23) FOLLOWER: the stony desert is below the start ?
GIVER: yeah , right , and we're going left from that ... and kind of round ... ... round left ... past the stony desert and down ... until you're kind of level with the bottom of the stony desert .
FOLLOWER: right , so i– ... i'm going left ?
GIVER: uh-huh .
FOLLOWER: and how far left ?
GIVER (PDU): left from the start and y– just ... just kind of round the outside ... of the desert ... you know right round the outline ... and then ... keep going down until you're kind of level ... with the stony desert.
FOLLOWER (CDU): but just round the outline of the desert?

Analysis:

1. PDU: assert[left from the start, right round the outline and then keep going down until you're level with the stony dessert]
   CDU: question[just round the outline of the desert]
2. CDU accepts PDU's instructions but requests clarification on one part
3. This example seems to be a clarification.

No new information or perspectives were raised here, neither does this case involve disagreement or rejection. Therefore it appears to be clarification rather than concession or DofE or correction. This shows that clarifications (which are by nature cross-turn relations) can also be signalled by "but". We will discuss this more in our conclusions from the evaluation.

- AMEX Tape 14:

(8.24) A (PDU): ok and that would be United fi ve seven four on the twenty fi fth of June
B (CDU): and then the return is another story we need to get them back the very latest

> Tuesday by noon but preferably they take an evening flight back, so I guess the airline depends on what we can get back conveniently

Analysis:

1. PDU: assert[that would be United 574 on the 25th of June]
   CDU: assert[we need to get them back the very latest Tuesday by noon but preferably they take an evening flight back]
2. CDU implicitly acknowledges PDU and doesn't disagree with anything
3. CDU introduces more constraints and a precondition on the flight mentioned in PDU
4. This is single-speaker "but" since it does not relate to anything in the previous turn but simply relates clauses in CDU.

Although CDU's precondition about flights that are convenient given the constraint of taking an evening flight probably also relates to PDU's proposal of a specific flight, this is not the "but" clause but arguably the next assertion made in CDU (i.e., "so I guess the airline depends on what we can get back conveniently") which responds to PDU. CDU gives the latest possible acceptable option (Tuesday by noon) and then argues that an earlier flight (assuming the "evening flight" refers to Monday evening) is preferable. This is why this is analysed as intra-turn "but".

8.2.2 Drawing Some Conclusions From the Data

We looked at a total of twenty-one examples, out of which we had three corrections, fourteen concessions, two cases with two possible interpretations both of which were DofE and concession, one single speaker (intra-turn) case and one clarification. While these numbers are far too small to draw any significant conclusions from, they do present a perspective of how the mechanisms presented in the thesis distinguish the three relations addressed on unseen examples.

Table 8.1: Acceptance & Rejection Features in the Evaluation

Relation	ExpAcc	ExpAck	ImpAcc	ImpAck	ExpRej	ImpRej
Conc.	8	1	5	3		
Corr.					2	1
DofE&Conc.	1		1			

Table 8.2: Characteristics of Relations in the Evaluation

Relation	Explicit Info	Implicit Info	Alternative	Explanation
Conc.	8	6	8	
Corr.	2	1	1	2
DofE&Conc.	1	1		

In Table 8.1, we summarise the acceptance/acknowledgement/rejection features of the relations in terms of whether they express explicit or implicit ("Exp" or "Imp" in the table categories) acceptance ("Acc"), acknowledgement ("Ack") or rejection ("Rej"). By explicit and implicit acceptance/rejection here we mean whether the acceptance etc. was made explicit or not. In contrast, Table 8.2 shows the distribution of cases which addressed implicitly vs. explicitly communicated information. For example, if an example refutes the answerhood of PDU, it refutes implicitly communicated information while if it accepts the assertion in PDU it is classified in the explicit information category. If it both accepts explicit information and rejects implicit information, it is classified as implicit, assuming that what is (in this case) corrected here is implicit. We also have categories for alternatives presented ("Alt") and explanations. In the case of concession, implicit and explicit information only relates to acceptance features, so this does not introduce new information. This is because concession by nature involves relating to a TC, an inferred relation which is inherently implicit. For correction on the other hand, the distinction between disagreeing with explicitly and implicitly communicated information becomes not only possible but necessary to distinguish the relation. Recall that the decision tree branched on disagreeing with implicitly communicated information to distinguish between concession and correction. For DofE, this distinction between explicitly and implicitly communicated information again becomes moot, as distinguishing DofE does not depend on this feature.

Certainly concession seems to involve both acceptance and the presentation of alternatives. There are not really enough examples of the other two cases (correction and the two examples which were ambiguous between concession and DofE) to draw any conclusions at all. As expected, in terms of the agreement feature, concession and correction involve mutually exclusive partitions of acceptance/acknowledgement for the former and rejection for the latter. DofE (of which there were only two examples, and these ambiguous between DofE and concession) did not involve any disagreement cases, though this is not surprising given the small size of the evaluation set. Regarding our earlier point, note that the fact that concession and correction partition the acceptance/rejection space does not discount the possibility that concession can involve acceptance of what is explicitly communicated while disagreeing with implicitly communicated information, as seen in cases involving supporting and rejecting relations X and Y w.r.t. a TC. We did not see any such cases in this evaluation, but that is hardly surprising given the small number of examples evaluated.

One point which should be considered when drawing conclusions is that the TC had to be predicted for all the concessive examples. Similarly, without a model of the speakers' beliefs, predicting correction and DofE relies on making educated guesses based on the context of the local dialogue context, a fact which must be kept in mind when considering this analysis. For correction and DofE this did not prove to be a serious drawback, since cross-turn correction simply involves the corrector rejecting something explicitly in the previous turn while either presenting an alternative or explaining why. In Example 8.18 the Follower corrects the Giver's acceptance of his location by explaining why his location should not be accepted.

Likewise, DofE also benefits from mental models, but can be reasonably predicted without access to them. Recall that predicting DofE involves assuming a defeasible rule with PDU as the lefthand side of the rule and the denied CDU as the right-hand side. For the purpose of this analysis, the rule arises from the pair of turns surrounding the "but". Looking back to Example 8.22, we see an example of this in the analysis, where determining the consequent of the expectation [always experiencing good leave policies at work] $>$ [there's always been good leave policies at work] requires resolving the pronominal anaphora in PDU. Resolving references, we get PDU: "i guess i

take [good leave policies] for granted because [the working situation] has always been that way but i know", and then the rule arises from the turns. So despite the lack of verification of our predictions for DofE and correction with speakers' mental models, we argue that it is reasonable to assume that the relevant beliefs hold for the purpose of this analysis.

One interesting issue that arises is whether we would have found the same TC as the one we predicted if we had taken more dialogue history into account in some of these cases, leading to the question of how much prior context is sufficient. In practice, we resolved the TC locally based on the pair of turns surrounding the "but" or including at most three prior turns, since the topic of discussion (i.e., TC) being discussed tended to be very locally relevant. Example 8.6 above illustrates that very little prior context is needed to determine the TC. In Example 8.15, while CDU refers to landmarks whose orientation depends partially on the dialogue history and their maps, the TC can also be analysed as a more local topic being debated in the PDU, CDU pair of turns, i.e., how to go from where the Follower is, either down and across and up, or diagonally across and down. So while the best analysis would of course take into account the full dialogue history in order to orient the speakers, our point here is that we have not missed out on too much by simply looking at the local context surrounding the related pair of turns, since we get a reasonable TC and relations X and Y that make sense in this context and shed light on what is being communicated in these cases.

In order to provide some validation for the procedures presented in this thesis, we showed that the decision tree distinguishing DofE, concession and correction relations worked as intended on these twenty-one unseen examples. That is, the theory predicted novel information (along with postulating that the relation in question held) which has utility in dialogue modelling. Since the decision tree derives from the distinctions made in the thesis, the evaluation reflects that these distinctions are reasonable, in the sense that they partitioned the examples as postulated and provided additional interpretation for a dialogue system attempting to make sense of cross-turn "but". That is, the associated expectation which is denied in DofE, the relations X, Y and the TC in concession and what is rejected and what is asserted (either as an explanation for the reason behind what was rejected or a replacement assertion) for correction all provide

important information for the hearer of the "but" to take account of when deliberating over how to respond. While we did not go into deliberation over responses in any depth in this thesis, we did present examples of the utility of modelling all three relations in terms of how it helps deliberation; for example, recall the Greta Garbo example. The hearer can respond to "But she never married" with "But beautiful people don't have to marry" if she recognises the underlying expectation which is denied and disagrees.

In the twenty-one examples considered, there were only two which were not covered by the procedures outlined in the thesis (these are the two that appear in the unclassified examples subsection at the end of the evaluation). One was an intra-turn single-speaker "but", and the other was a clarification. Clarifications are by nature cross-speaker relations, and as we see can involve "but", but if we consider the example again (reprinted below)

(8.25) GIVER (PDU): left from the start and y– just ... just kind of round the outside ... of the desert ... you know right round the outline ... and then ... keep going down until you're kind of level ... with the stony desert.
FOLLOWER (CDU): but just round the outline of the desert?

we can see that there is nothing in contrast besides the meta-level lack of understanding that the Follower has with the accepted instructions. Clarifications do not involve contrast as such and were omitted for that reason from this thesis study.

We did not find any cases which were covered which should not have been, though presumably one needs to look at a significant amount of data to find such cases. The fact that clarification was not covered is a promising indication of this.

There were also not as many ambiguous cases as imagined. Of the two cases that were ambiguous (Examples 8.21 and 8.22), both were analysed as DofE and concession. In order for the DofE interpretation to arise, $PDU \rightarrow \neg CDU$. This means (minus verification via mental models) that this rule has to be reasonable given PDU and CDU. Recall the analysis for Example 8.22 (abbreviated below),

- DofE: [always experiencing good leave policies at work] $>$ [there's always been good leave policies at work]

- concession: TC=[there's always been good leave policies at work] X=supporting-evidence(PDU,TC); Y=counter-evidence(CDU,TC)

note the similarity between the right hand side of the rule (i.e., the negated CDU) and the TC. They are basically identical. Consider the analysis for Example 8.21 below

- DofE: [right now I'm working within your reservation] > [necessary information to finalise reservation is present]

- concession: TC=[reservation is nearly finalised]; X=provide-evidence(PDU,TC); Y=introduce-new-precondition(CDU,TC)

here the only difference is that the TC omits mention of the "necessary information". This implies that one interpretation or the other may be redundant and the relation definitions may overlap, except that in both examples the concessive analysis introduces new information about how PDU and CDU are related to the TC. This begs the question whether DofE is redundant. The main differences between DofE and concession are (1) that the denied expectation arises from the speaker's beliefs in DofE while the TC arises from the dialogue history, and (2) that DofE can involve disagreement with the PDU, where CDU denies a reasonable expectation arising from PDU in order to provide evidence against PDU, while concession involves minimally acknowledgement of PDU. These differences are not negligible, and we feel that they each introduce novel and complementary information w.r.t. the cross-turn "but"-cued relation being communicated. Furthermore if we had access to speakers' mental models to confirm that the relevant expectations held, we might have uncovered different denied expectations for DofE than the ones we found, thereby further enriching the DofE interpretation with novel complementary information with respect to the concessive interpretation.

Summary

We evaluated the proposed theory for distinguishing DofE, concession and correction on unseen data and found that it covered all the data except for clarifications, which were not addressed by the thesis from the outset. Furthermore, it did not misclassify any examples that should not have been classified as such. Therefore the procedures presented cover the data and work as intended. Given that the main goal of this thesis is to provide interpretations for cross-turn "but"-cued relations which give useful

information for modelling what was communicated and deliberating what to say next, we feel that this analysis shows that the interpretation procedures presented have done that, since in all examples analysed, they shed new and salient information which has immediate utility.

8.3 Implementation

We now go on to present the DofE in NTOD procedure which was implemented in a working model of dialogue. We will start by discussing the EDIS model of dialogue, since this is the model in which our implementation was conducted. We will then present our implementation and indicate how it ties into the IS approach.

8.3.1 The EDIS System

The EDIS system is a version of the PTT model of dialogue which focuses specifically on how obligations are addressed and how information is added to the common ground. Our implemented structure follows the PTT system closely but makes some simplifying assumptions. The EDIS IS is illustrated in Figure 8.3.1.

Both speakers (Wizard and Caller) are represented in the EDIS IS although only the Caller's intentions are represented as they do not (at least at the time of writing the TRINDI manual) model misunderstandings.

We propose a set of update rules operating on multiple levels of the IS in order to model the "but" triggered relations put forward in this thesis. Propositions are represented as simple functions of the predicate (e.g., "Bob eats bagels" becomes $eats(Bob, bagels)$) and we focus more on how update will work on the "but" cases we study rather than on representing utterances formally in a particular formal semantic theory.

Our implementation adds a Task Planning field to the IS to address TOD involving speakers' plans. Task Plan History (TPH) is analogous to Dialogue History (DH) and holds the most recent Task Actions (TAs are similar to Conversational Actions in the DH). For example, TAs can represent plans proposed, preconditions, effects, etc. that are communicated in a given turn (DU). We assume a TA interpreter that interprets

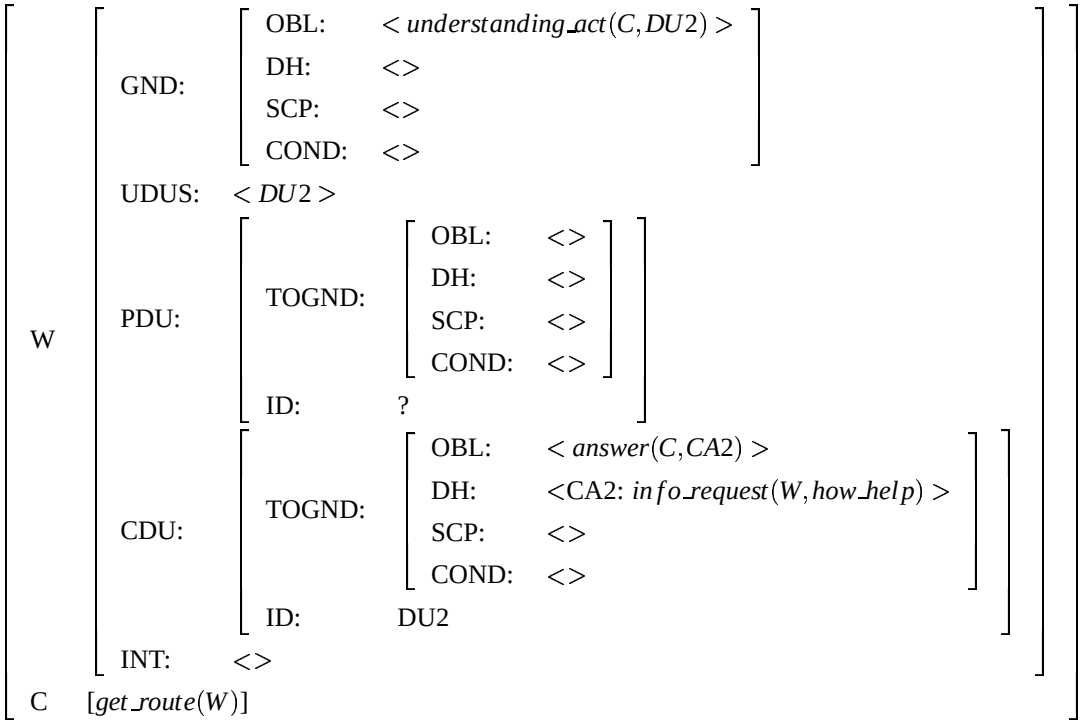

Figure 8.3: Information State Structure in EDIS System for “how can I help”

Conversational Actions (CAs) and determines what TAs are being communicated by accessing the speaker’s task-plan. The TA interpreter should be called from an update rule as soon as the CAs are interpreted and the IS has been updated with the CAs communicated in the current turn. The TAs returned by the TA interpreter should then by added to the IS’ CDU.TPH field. The specifics of when the procedures outlined in the thesis called the CA and TA interpreters, what the arguments were and what was returned by the interpreters (in the form of IS updates) was specified more precisely in the sections outlining DofE, concession and correction earlier in the thesis.

In discussing both CA and TA intepreters, we address both TOD and NTOD in the belief that the two are often intermingled, and many of the same mechanisms can be brought to bear upon both types of dialogues. Most dialogue systems tend to address only TOD and NTOD has been eschewed mostly because coherence is so much less focused on a particular topic in comparison with TOD. We will not claim that we address the many different types of NTOD behaviour, but rather argue that while we cannot address all the different factors involved in NTOD, we can still make use of

some of the interpretation machinery in place for TOD that to produce appropriate IS updates for NTOD.

Certainly the EDIS and PTT frameworks are simply that, frameworks in which many different sorts of dialogue phenomena can be modelled. The update rules operate in stages in EDIS3, a version of the EDIS system (TRINDI, 2001), and all rules in each stage and each stage itself operate sequentially. We used the TRINDI toolkit for our implementation, in which the update rules are generally written in prolog. We will now sequentially go through the general update procedure for EDIS3, which is included in the Appendix. The particular update procedure described here is specific to EDIS3. Different IS models have different approaches, as the IS is simply a framework wherein dialogue modellers can implement their theories. In the first stage the contents of PDU get pushed into UDU and the contents of the previous utterance (still in CDU) is pushed into PDU and a new CDU is initialised. In the next stage backwards grounding acts are performed, so if an acknowledgement exists in the *latest_moves* field, PDU is merged with GND (Ground), and the acknowledgement is recorded in GND.DH. Then all other backwards and forwards acts are performed, for example, asserts, information requests, agreement, checks, etc., and they all require an understanding act on the part of the hearer. This is the stage at which the rules incorporated to address cross-turn "but" are placed, i.e., all these update rules are of ruletype 3. At the next stage (ruletype 4), obligation resolutions are handled (Matheson *et al.*, 2000). Then implicational acts are addressed. These include acts of the form $accept \rightarrow scp$ or $agree \rightarrow scp$, where *scp* stands for "socially committed proposition" and indicates that speakers have understood and accepted the truth of what was asserted (etc.) formerly and both speakers are now committed to the truth of this information. The last ruletype addresses intention-handling rules. This means that if obligations exist, e.g., to perform an understanding act, the intention to perform the act is adopted. Similarly, the intention to accept a directive or an assert based on the obligation to address it is pushed onto the INT (intentions) field. The last rules hand control to the deliberation component and determines if intentions imply other intentions or sub-intentions and push these onto INT also.

8.3.2 Our DofE Update Rule

The following rule updates the IS for the simplest type of “but”-triggered DofE, the NTOD case presented in 5.1.1. Rules in PTT are implemented as a series of conditions and effects; if the conditions are matched, the effect is produced. The effect of the following rule is to push a condition saying that if the other Dialogue Participant (DP, i.e., the hearer of the DofE) accepts the DofE, then the rule (i.e., the defeasible expectation) becomes a socially committed proposition:

```
%  effect:  push(COND,accept(o(DP),ID) -> scp(o(DP),RULE)) (modelled
%           after assert2 urule)
```

The DofE update rule starts by initialising variables and tests whether the latest move has a DofE predicate. We discuss how the DofE predicate is added to the IS after presenting the outlines of the rule.

```
urule(doDofe, ruletype3,
[in(latest_moves, Mv),
 val(hearer,H),
 Mv::val#rec(atype,Move),
 Move::val#rec(pred,dofe),
```

The next bit of code gets the values of the previous turn (PDU). Notice that this case requires that the previous turn is an assertion.

```
 in#rec(w^pdu^tognd^dh,ACT),
 ACT::val#rec(atype,ATYPE),
 ATYPE::val#rec(pred,assert),
 ACT::val#rec(id,ACTID),
%get proposition and arguments
 ATYPE::val#rec(prop,APROP),
 APROP::val#rec(pred,APRED),
 APROP::val#rec(args,ARGS),
```

Now we get the values of the current turn, which in this case also needs to be an assertion, just as in the algorithm in 5.1.1, which left addressing questions and commands for future work.

```
%get CDU assertion CA
 in#rec(wˆcduˆtogndˆdh,CURACT),
 CURACT::val#rec(atype,CURATYPE),
 CURATYPE::val#rec(pred,assert),
 CURACT::val#rec(id,CURACTID),
```

Then variables are assigned to the predicate, argument and proposition in CDU in order to simplify addressing them later.

```
%get proposition and arguments
 CURATYPE::val#rec(prop,CURPROP),
 CURPROP::val#rec(pred,CURPROPPRED),
 CURPROP::val#rec(args,CURARGS),
%get current ID
 val(next_dh_id,HID),
 print_util::next_dh_idname(HID,HIDname),
```

The line below calls the actual procedure used to distinguish and assign DofE to the update. We will give the code implementing this procedure after finishing the description of this update rule.

```
dofe_alg::dofe(APRED,ARGS,CURPROPPRED,CURARGS,Rule)],
```

This line ends the conditions of the update rule, and the next bit of code describes the effects of the rule. The code below updates CDU.TOGND.DH with the DofE relation if the *dofe_alg* procedure succeeds. (If *dofe_alg* fails, the rule will fail and the effects below will not be updated.)

```
[increase(update_cycles), increase(next_dh_id),
push#rec(wˆcduˆtogndˆdh,record([atype=
   record([pred=dofe,
           dp=c,
           args=stackset([ACTID,CURACTID, Rule])]),
           clevel=2,
           id=HIDname])),
```

The next bit of code pushes obligations onto the hearer (W, i.e., W.CDU.TOGND. COND) so that if W accepts the defeasible rule communicated by C, both speakers

are socially committed to this rule (i.e., it is grounded). That is, we follow Matheson *et al.* (2000) in using the CONDitional attribute, assuming that assert acts and also now DofE and the expectation it denies (and subsequently other discourse relations) impose obligations on the hearer to respond to them. Once something is accepted, whether it is what is asserted, the DofE relation or the rule/expectation, this then becomes something that both speakers are socially committed to. We show the conditional rule for the case in which the DofE rule is accepted. The lines below basically say that a condition is pushed into W.CDU.TOGND.COND such that $accept\_rule > scp(dofe\_rule)$.

```
push#rec(wˆcduˆtogndˆcond,
   record([atype1=
      record([pred=accept_rule,
              dp=H,
              args=stackset([HIDname])]),
           atype2=record([pred=scp,
                          dp=H,
                          prop=record([pred=dofe,
                                       args=stackset([Rule])])])]))
```

Of course other conditions are imposed for DofE; the denial of the rule may also be accepted, in which case both the rule and its denial are accepted by the hearer. The condition above shows the case in which only the rule was accepted. Currently, just the assertion can be accepted by the hearer of the DofE, which would be represented by a standard *assert_accept* as originally proposed in Matheson *et al.* (2000).

DofE procedure

Now we will present the $dofe\_alg$ algorithm called in the DofE update rule above. Recall that the rule was of the form *antecedent* > *consequent*, where the antecedent came from the assertion in PDU and the consequent was the inverted/negated assertion in CDU, (i.e., the variables CURPROPPRED or NewRHS after it is negated). The Greta Garbo example has “But she never married” in CDU, and “she married” after negation. Once this is generalised (i.e., uninstantiated), it becomes $marry(X)$. The antecedent or left-hand side (LHS) of the rule is bound to the variable APRED below, and after generalising it becomes $beautiful(X)$ or in this case, simply “beautiful people”.

The parsing of utterances is currently canned, as the point here is to illustrate what needs to be added in order to interpret discourse relations. We assume that utterances are parsed into a logical representation in which changing constants to variables (e.g., $beautiful(greta)$ to $beautiful(X)$ is straightforward).

```
dofe(APRED,ARGS,CURPROPPRED,CURARGS,Rule):-
RHS=CURPROPPRED,
NewRHS=marry,
LHS=APRED,
NewLHS=beautiful_people,
Rule=[NewLHS,NewRHS].
```

Of course this is the simplest of relations modelled in this thesis, but it is our hope that it illustrates the feasibility of implementing the rest. In a working system with speech recogniser, parser, planner, etc., procedures like this should be perfectly feasible. In some cases, like concession in TOD, a TA interpreter which determines what task actions are being communicated (given PDU's CAs and the task plan) will be necessary. Once CA and TA interpreters are available, the system should be able to interpret expectations given speakers' beliefs and task plans rather than rely on canned rules.

Summary

The role of this implementation is simply to show how straightforward it is to incorporate the procedures presented throughout this thesis. As such it is not intended to describe a fully implemented system, but rather to give simple evidence of the ease of implementing the ideas discussed in the thesis.

This chapter began by summarising and distinguishing DofE, concession and correction, determining their characteristics and how they differ. We then presented a decision structure which distinguishes the relations according to the theory proposed in the thesis. We went on to analyse the theory based on this decision structure on unseen examples to evaluate its performance and coverage. We finished by showing how one of these procedures was implemented in a working model of dialogue to indicate the feasibility of our approach. In the next and final chapter, we consider unsolved issues and areas for future research.

Chapter 9

Looking Forward

"The word is the virus." William S. Burroughs

A single word can often be polysemous, and it can have countless meanings which are context-dependent. In the case of "but", we have seen that it can correct prior turns, deny speakers' expectations, and trigger the interpretation of expectations in argumentation, where the range of expectations are bound only by the situations and ideas which the speakers choose to speak about. However numerous (synchronous) interpretations speaks of meaning proliferation which is an undesirable feature of linguistic interpretation systems. The role of pragmatic mechanisms such as the ones proposed in this thesis act to limit meaning proliferation by introducing context-specific constraints on interpretation and identifying generalising mechanisms underlying broad groupings of meaning.

The central purpose behind this thesis has been driven by the need for an account for cross-turn discourse relations in dialogue, to which end we focused the study on cross-turn "but"-cued relations. In this thesis, we have taken the perspective that the interpretation of spoken language is subjective, and tied to the context in which the Speech Actions occur. This context can include both the speaker's external environments and their own mental model of the world, the latter of which often involves beliefs, expectations, plans and goals which can differ between speakers.

In choosing to focus on how speakers interpret relations between utterances across speaker-turns, we modelled these interpretation and deliberation processes from a subjective perspective, allowing for misunderstandings and differences between speaker's mental models. We presented a mechanism for interpreting what is communicated with these cross-turn relations, modelling them in a framework that enables recognition of the differences between speakers' mental models.

For the most part, multiple interpretations do not hold in this approach, as the three "but"-cued relations addressed focus on different aspects of meaning which are often incompatible. In a small number of cases two interpretations may hold, but we argued that these cases were only enhanced by enabling both perspectives of what was being communicated, which were not mutually exclusive alternatives, but rather two aspects of meaning which can both simultaneously be conveyed and reflect the richness of communication content in these cases. We saw a few examples of this in the last chapter, where it was also argued that the two relations are not subsumed by each other, but rather reflect different but complementary information about what was communicated. The three relations addressed in this thesis generalised over the data to form three distinct and informative ways of interpreting a wide range of cross-turn "but"-cued relations between turns.

We will now go on to address areas for future work. We will begin by addressing how the interpretations of DofE, concession and correction can be used to aid deliberation of responses and discuss in more detail what one might need to take into account to deliberate about responding to these relations. We will then briefly discuss the generation of cross-turn "but"-cued relations themselves before finally addressing the larger issue of how this approach might lend itself to the analysis and interpretation of other cross-turn discourse relations cued by discourse markers.

9.1 Deliberating Responses to Cross-Turn "But"

In this section we will briefly discuss ways in which a speaker might respond to cross-turn "but" which take advantage of the interpretation of cross-turn "but" relations. Here we will simply present salient information that the hearer of a "but"-signalled

relation might want to deliberate to respond to the "but"-turn given the interpretation presented in the thesis. The basic motivation is to make use of the salient information about the speaker's expectations, intentions and goals revealed and communicated in the interpretations discussed thus far to aid the hearer of the "but"-signalled relation to respond appropriately. While he might not respond immediately to the inferred discourse relation communicated, he is socially obliged (Kreutel and Matheson, 2000) to respond if he has relevant conflicting beliefs, goals or expectations with respect to the previous turn in order to prevent future misunderstandings and disagreements. Although we will not discuss how speakers choose what to express next among the many things they deliberate over, one of the things they may want to communicate is disagreement with inferred relations and their accompanying inferences.

9.1.1 Deliberating Responses to DofE

We will begin by considering how information gained from interpreting DofE can be used to deliberate over salient responses to either the DofE, its accompanying (and denied) expectation, or the "but"-fronted assertion itself. Recalling the treatment of DofE presented in Chapter 5, the hearer of the DofE can disagree with the expectation or rule underlying the DofE, the assertion communicating this rule's denial (i.e., the DofE), both assertion and rule or neither, based on her own beliefs. If she disagrees with some or all of what is communicated, she is then socially obliged to respond and indicate this, thereby preventing the speaker of the DofE from assuming that she agrees with him.

We consider below the different ways in which the hearer (A) might respond to the speaker of the DofE B's turn depending on her (A's) private beliefs. We frame these different possible situations in terms of how they might be distinguished given an IS (Information State) in which DofE has been updated in CDU (Current Dialogue Unit; i.e., the "but" turn) and the hearer's (A's) turn is next.

1. If CDU.DH contains a $dofe$ CA (uttered by B), then

    (a) if A disagrees with both the rule and B's assertion, then neither must be in

his private beliefs[1] or for TOD in his Task Beliefs (TB) or in the grounded TPH, and additionally, he might have conflicting TB or plans.

(b) if A agrees with B's assertion only, then the assertion needs to be in A's private/task beliefs, and possibly also information which conflicts with the rule/plan.

(c) if A agrees with the rule only, then the rule itself needs to be in A's private/task beliefs, and possibly also information which conflicts with the assertion in his private/task beliefs or plan.

(d) if A agrees with the instantiated DofE relation holding, then the rule and the instantiated propositions all need to be in A's private/task beliefs.

The motivation for responding swiftly to contradicting beliefs is even more immediately important in TOD, where correcting misunderstandings and disagreements fast can bypass much wasted time spent afterwards trying to determine what is in conflict in order to decide upon a plan that both speakers want. Upon hearing B's DofE, A must then respond appropriately. If A also infers the nature of B's denied expectation, this can lead to much more responsive deliberation.

9.1.2 Deliberating Responses to Concession

In Chapter 6 we interpreted concessive "but"s in both TOD and NTOD, and we showed how the IS is updated in these cases. We presented two basic schemes, both of which involved relations *X* and *Y* between the TC (i.e., salient claim under discussion) and the PDU and CDU (respectively); one scheme involved an additional relation *Z* between the propositions in PDU and CDU. We distinguished pairings of *X* and *Y* predicates corresponding to a range of concessive situations. In all of these cases, we found that the speaker of the concessive "but" turn (CDU) both (1) agreed, accepted, or (minimally) acknowledged the former turn (PDU) while (2) presenting a preferred alternative argument w.r.t. the TC.

[1] Alternatively, A might have in his private beliefs either the negated rule and assertion or information which is incompatible with the rule and assertion.

So what do we need to consider when we know that the previous turn responded to the preceding dialogue concessively, and we know the issues being discussed and what is agreed to, objected to, etc? Well, if the hearer (A) of the concessive response interprets the response concessively, she may agree or disagree with any of the information conveyed by the speaker (B) in his concessive turn or find what he says irrelevant. Depending on what else is communicated (e.g., TC, stance, planning operator relationships, objections, agreement, etc.), A can (1) agree, (2) disagree, and do so (a) partially or (b) with everything communicated. A can also agree/disagree with (i) what is actually being communicated (explicitly) or (ii) with the relation (i.e., concession) or relationships implied, or (iii) with the relevance of what is being communicated. If A disagrees, she can do so by (I) presenting alternatives or (II) criticising what she disagrees with. A's response to B's concession will be determined by her own beliefs, the prior dialogue, and how B's utterance fits in with her own goals/beliefs. For example, if B's concession presents an alternative to A's proposal which she finds better in terms of achieving her goals, or if B's answer introduced new information that causes her to change her beliefs, she can also indicate this. We will illustrate below how A might respond given her own beliefs with respect to B's concessive turn.

- agree/disagree partially/completely
  - with explicit content/implicit content (e.g.,discourse relations, meta-level predicates like alternatives, etc.)
  - with relevance of what was communicated
  - if A disagrees, she can present alternatives/criticism
  - if A agrees, she can add supporting evidence, express agreement directly, or implicitly agree and continue the discussion (without addressing B's concession)

These are the primary ways in which A can directly address B's concession. Of course, if an implicit strategy of grounding is followed (TRINDI, 2001), then A can continue on with the discussion, not directly addressing B's concession However if A disagrees with the concession relation, then she should respond as soon as possible in

order to bypass further misunderstandings in the future, and so we present the above, fairly obvious, ways in which A can address the concession relation.

Furthermore, of the many aspects of the concession that A can respond to, she can present her perspective w.r.t. implicit information like the concession relation itself, or she could address *Y* or comment on what B said explicitly either by presenting alternatives or by criticism (in the case that she disagrees with something). Furthermore, A could criticise the relevance of *Y* or of what B communicated explicitly.

Deliberation should simply involve testing the various agreement features in each case. This involves determining for concession whether the TC is agreed with, whether the preceding speaker's expectation w.r.t. the TC is OK and whether the preceding speaker's assertion is reasonable given the responder's IS. Also, *Y* should be responded to; for example, the hearer needs to determine whether an alternative (if presented) is also an alternative for him (i.e., whether the alternative is valid, whether the relevant options are valid singly, and whether the hearer also sees them as alternatives), whether criticism and alternative beliefs or opinions are agreed with, and lastly whether invalid actions are also seen as such, and whether they are relevant.

9.1.3 Deliberating Responses to Correction

We will now focus on considerations for deliberating a response to correction, given the updates attributed to correction cases. In Chapter 7 we saw that the constellation of CAs involved in corrections all involve at least *assert*, *disagree*, and *correct* CAs. Corrections of answers to *why*-questions involved *reject* instead of *disagree* and *accept* instead of *assert*; here the corrector accepted the assertion but not that it works as an answer to the question, so its answerhood is rejected. If the assertion is not accepted, there is no *accept* CA and only an *implicit_disagree* CA. Also, the type of CA which is corrected in the *correct* CA will give a lot of useful information about what the corrector believes so that the hearer can respond appropriately, depending on how her own beliefs relate to what she infers about the corrector. We will assume here that the most informative information available in the case of a correction will be found in the *correct* CA itself, since this indicates which CA is being corrected, and further information as to why or providing an alternative will be found in an *assert* CA along

with the *correct* CA. The focus on just three CAs greatly simplifies deliberation of a response to correction.

Now let us consider the various facets of response one can make to a *correct* $(Speaker, CA_b)$ CA. In some cases corrections can hold with part of a CA or part of the content of an assertion, etc. Corrections can also hold with planning operators and question types. Certainly the features of agreement and disagreement, e.g., partiality, content (explicit or implicit, relevance), and disagreement mechanisms (e.g., presenting alternatives or criticism) also apply to corrections. Recall also that the assertion accompanying a *correct* CA either states what is wrong about the CA being corrected or introduces alternative information that the corrector considers more relevant. Both these situations could involve the introduction of new information to the correctee or highlight common knowledge and illustrate its relevance in the given situation. In any case, the correctee must interpret what is being corrected and then respond, either by accepting that the corrector is right, or by disagreeing and explaining why.

Depending on what is corrected, the correctee needs to determine his own standpoint w.r.t. the correction by checking his beliefs, via the appropriate fields of the IS representing his beliefs, plans, what he has said, etc. Deciding to accept correction involves comparing the new (corrected) information with his original views in order to determine if the new information is compatible or not given his own beliefs and reasoning. While it is possible for a speaker to accept a correction and revise his beliefs, he should also be able to disagree if he knows of something wrong or incompatible in the corrected information. Recall also that corrections can involve corrections of implicitly communicated information like discourse relations and inferred expectations as well as with assertions, commands and questions. Corrections of answers behave similarly, as they involve correcting the answerhood rather than the explicit assertion communicated.

So to summarise, one can respond to what is asserted explicitly in the correction or to implicitly communicated information like discourse relations. An answer can be corrected by introducing a necessary precondition that needed to be satisfied before the answer was acceptable. One can also have corrections of mistakes in slot-values in questions which show how the question is invalid. In almost all of these cases, the IS

was simply updated with the more or less static constellation of *disagree*, *assert* and *correct* CAs.

Responding to "but"-signalled correction requires first comparing what is corrected and asserted by the corrector with the beliefs of the correctee in order to update the IS with a list of the resulting conflicting information found in the speaker's beliefs or in the dialogue hhistory (since the correctee might have previously uttered conflicting information). The conflicts should then be turned into intentions to assert them. After deliberating over which of these intentions should actually be expressed next, these assertions of conflict should be expressed. A benefit of this approach is that conflicts (raised by corrections in this case) are always straightened out without delay, hopefully reducing the number of misassumptions which might otherwise be the case.

Summary:

In this section we have focused on how hearers of the cross-turn discourse relations presented in this thesis might respond (depending on their own beliefs), in the hope that by doing so, they avoid future misunderstandings and disagreements. The focus on deliberation of responses to cross-turn relations in this chapter is meant to serve as a means of indicating the utility of the approach laid out in this thesis as well as providing guidelines for how to implement these considerations in a working dialogue's deliberation process.

The evidence that similar techniques can be used to facilitate deliberation of appropriate responses to DofE, concession and correction is especially significant because it points to the universality of this approach. Since cross-speaker relations involve speakers conducting a coherent dialogue, this work serves as a step towards facilitating the modelling and deliberation (of one side) of a coherent dialogue.

9.2 Generating Cross-Turn "But"-Cued Relations

Another area of work that arises from this thesis addresses the generation of "but"-cued turns themselves. This discussion follows directly from the features involved in distinguishing the relations. That is, the basic features of DofE, concession and correction which lend themselves to the interpretation of these relations also lend themselves to

the generation of the relations themselves. Since generation also involves weighing the information gleaned from interpretation with the speaker's larger communicative goals and plans, much of this is beyond the scope of this work[2]. However we can specify elements that need to be present in the deliberation process for a speaker to express cross-turn "but"-cued relations.

In the interpretation, we have specified necessary features for identifying various relations and we argue that similar information might very well be useful for generating the relation itself. For example, the belief in a salient expectation and knowledge of a fact contradicting its consequent can enable the generation of DofE. Consider the case of concession: salient information about alternatives with respect to the previous turn which also coincides with acknowledgement of this previous turn might enable concession. If B rejects something which A uttered and knows of either an alternative or an explanation, this enables generation of correction.

Of course speakers will have other information to deliberate over expressing as well as higher-level communication goals to be expressed, and these only add to the range of possible information which might be expressed next. Although this is not a topic which has been addressed in this thesis, it is an important and relevant one. Like deliberating responses to cross-turn "but"-cued relations, generating cross-turn "but" itself arises directly from the interpretation provided by this thesis, and it is included here as an area for further research and implementation in deliberation models.

9.3 Extending the Underlying Mechanism

These different aspects of interpreting and deliberating over responses to cross-turn "but" relations lead to a larger and more central question. How can the interpretion of cross-turn "but" relations be extended to account for other cross-turn relations? In this subsection we will abstract away from the actual steps involved in the interpretation of these relations and focus instead on the general procedures involved to obtain the relations' interpretations. The goal is that the general mechanisms involved here may apply to other cued cross-turn relations' interpretations as well. The idea is that given

[2]See Green and Carberry (1992) for an approach involving generation of indirect replies which convey conversational implicatures and also Hovy (1993) for approaches in related areas.

the more general case below

(2) A: X

B: < *cue* > Y

we have a basic tool in the form of a simple procedure for determining how X and Y are related. This procedure of course does not take into account the semantics of < *cue* > or any intuitions about how < *cue* > relates X and Y, and this must be added by the researcher investigating the given cue. However it is possible to sketch out a basic procedure for investigating the interpretation of the cue in question. Note that this approach will only work for cues which are medial in monologue, since we consider turn-initial < *cue* > constructions here. Sentence-initial cues in monologue cannot relate material across the turn grammatically. So the basic steps for interpreting medial cues are as follows, assuming an IS model of dialogue here (though of course this is not necessary):

1. If you see a turn-initial (or approximately turn-initial) < *cue* >, then
   (a) Determine how the previous turn is grounded by the current speaker; e.g., is it agreed with, disagreed with, clarified, etc.
   (b) Check whether the part of the utterance that immediately follows < *cue* > appears in the IS in dialogue history or in beliefs or intentions and if it appears as stated or if it is the argument of some function expressing (for example) agreement etc.
   (c) Search the IS for the speaker of the < *cue* >'s (let's call her B) beliefs and determine whether there are any rules in beliefs which have either the current or preceding turn as either antecedent or consequent; it is possible that the IS will contain an inference made from the turn.
2. After updating a < *cue* > turn, consider how the hearer's beliefs, dialogue history and intentions relate with the updated IS and deliberate responding to these issues first.

Of course some intuitions on the sort of relation conveyed by < *cue* > in monologue will help guide what sort of relations to start looking for, although (as was the

case for “but”) different relations are possible for dialogue that are impossible for monologue (e.g., disagreement, rejection, correction, etc.). It should also be pointed out that despite framing this procedure within the context of an IS model of dialogue, the basic ideas underlying the procedure should generalise to other models.

Although these steps only spell out where to look and very roughly what sorts of things to search for, it is a first step towards going about determining how two turns are related, in the hope that we can develop interpretation algorithms for other medial cues when they appear turn-initially in dialogue, thereby going some way towards addressing the initial question of explaining how two turns (when given a cue clue) are related in dialogue.

9.4 In Closing

This thesis started by asking why some responses seem more reasonable than others in dialogue, and we saw that context plays a crucial role in determining whether a response is reasonable or not. We described degrees of acceptability of a response as an indication of the degree to which the turns are coherent. We started by analysing cross-turn “but” and postulating some necessary criteria for coherence between turns in these cases. We then bore these hypotheses out by testing whether these criteria, when incorporated into the IS framework, rendered interpretation algorithms capable of filling in the missing information, e.g., denied expectations in DofE, which makes the dialogue coherent. In this chapter we finished by presenting possible areas for utilising this interpretation to determine how the hearer can then deliberate an appropriate response either to the “but” turn or to express a “but”-cued relation based on her own mental model and the dialogue history. Finally we presented the skeletal structure of investigation our approach took in terms of how other medial cues might be interpreted when they link material across turns in dialogue.

It is our hope that this work will help other researchers interested in developing systems capable of interpreting and responding to speakers coherently.

Appendix A

Implementational Details

Here we will present the code referred to in Chapter 8. First we will present the EDIS3 update rules used in our implementation, and then give the additional modules adapted to work with our modified set of update rules. Lastly we will present a concessive update rule.

A.1 EDIS3 Update Rules

As discussed in Chapter 9, the EDIS update rules are classified into various types in order to impose general orderings on when things happen - so ruletypes 1, 2, and 3 occur in order first, and all rules in each type are tried in turn. A complete ordering of all the rules would achieve this, of course, but not the final stages, in which deliberation happens, followed by obligation resolution and finally conditional resolution. So what actually happens is that types 1, 2, 3 and then 6 are processed, and then 4 and 5 are cycled through as specified in the update algorithm.

This version uses the COND attribute as proposed in Matheson *et al.* (2000), and also assumes that assert acts impose an obligation to respond on the hearer (by addressing them). We have added similar obligations to discourse relations. For assert acts, the deliberation stage turns these obligations directly into accepts, mimicking implicit acceptance, and the accepts discharge the relevant conditionals in COND.

```
/*****************************************************************************
```

```
         name: update_rules.pl
      version: Aug 1st, 2000
  description: The EDIS3 update rules
       author: Colin Matheson and David Traum

last modified by Kavita to model discourse relations across speakers
signalled by "but", 11/6/03.
*********************************************************************/

:- module(update_rules, [rule/3, rule_class/2]).

:- op(800, fx, ['!', not]).
:- op(850, xfx, ['$=', '$==', and, or] ).

/*-------------------------------------------------------------------
      Port to new rule syntax
-------------------------------------------------------------------*/

rule( RuleName, Preconds, Effects ):-
        urule( RuleName, _, Preconds, Effects ).

rule_class( RuleName, Class ):-
        urule( RuleName, Class, _, _ ).

/*-------------------------------------------------------------------
     Definitions of u-rules, syntax:
     urule( RuleName, RuleType, PrecondList, EffectsList )

     The basis for these rules is Figure 2 in the NAACL 2000 paper - where
     appropriate, the relevant rule from Figure 2 is provided.
-------------------------------------------------------------------*/
```

```
%%%  ---------------          RULETYPE 1          --------------------

%%%  Copy CDU contents to PDU, make a new CDU, push the ID into UDUS,
%%%  and increment the DU ID number.

urule( makeNewDUandCopyCDU, ruletype1,
       [ val#rec(w^cdu,WCDU),
         val(next_du_id,DU),
         print_util::next_du_idname(DU,DUname) ],
       [ set#rec(w^pdu,WCDU),
         set#rec(w^cdu, record([tognd=record([scp=stackset([]),
                                              obl=stackset([]),
                                              dh=stackset([]),
                                              cond=stackset([])]),
                                id=DUname])),
         push#rec(w^udus,DUname),
         increase(next_du_id) ]
     ).

%%%  ---------------          RULETYPE 2          --------------------

%%%  (\cite{TRINDI}, cf D1.1 p36, Update Rule 2).  Do W's Backwards Grounding
%%%  Acts. If latest_moves contains an acknowledgment, merge PDU and GND,
%%%  push the acknowledgment into GND^DH, and remove PDU from UDUs.

%%%  act:     ID:2, ack(DP,DU1)
%%%  effect:  peRec(w^gnd,w^pdu^tognd)
%%%  effect:  remove(DU1,UDUS)

urule( doBGActsW, ruletype2,
       [ val(latest_speaker,w),
         in(latest_moves,Mv),
         Mv::val#rec(atype,Move),
```

```
          Move :: val#rec(pred,ack),
          val#rec(w^pdu^id,PID),
          val(next_dh_id,ID),
          print_util::next_dh_idname(ID,IDname) ],
        [ increase(next_dh_id),
          increase(update_cycles),
          is :: pePathRec(w^gnd,w^pdu^tognd),
          push#rec(w^gnd^dh,
                   record([atype=Move,
                           clevel=2,
                           id=IDname])),
          delRec(w^udus,PID) ]
      ).

%%%% More or less the same, except that C's rule has to pick up the ID of the
%%%% act to be acknowledged - this should be done separately in deliberation
%%%% as it is for W.

%%%  act:     ID:2, ack(DP,DU1)
%%%  effect:  peRec(w^gnd,w^pdu^tognd)
%%%  effect:  remove(DU1,UDUS)

urule( doBGActsC, ruletype2,
        [ val(latest_speaker,c),
          in(latest_moves,Mv),
          Mv::val#rec(atype,Move),
          Move :: val#rec(pred,ack),
          in#rec(w^gnd^obl,OBL),
          OBL :: val#rec(pred,uact),
          OBL :: val#rec(dp,c),
          OBL :: val#rec(args,ARGS),
          val#rec(w^pdu^id,PID),
          val(next_dh_id,ID),
```

```
          print_util::next_dh_idname(ID,IDname) ],
        [ increase(next_dh_id),
          increase(update_cycles),
          is :: pePathRec(wˆgnd,wˆpduˆtognd),
          push#rec(wˆgndˆdh,
                   record([atype=record([pred=ack,
                                         dp=c,
                                         args=ARGS]),
                           clevel=2,
                           id=IDname])),
          delRec(wˆudus,PID) ]
      ).


%%%  ---------------          RULETYPE 3          --------------------

%%%  Do all other Backwards and Forwards acts.
%%%  The first four rules handle the fact that all rules of this type require
%%%  an understanding act on the part of the hearer.  Note that there has
%%%  to be a check that there actually is an appropriate core act in
%%%  latest_moves - simplified here as a test to see if there's an assert,
%%%  an agree, an info request, or a check.  Otherwise bare understanding
%%%  acts would result in an obligation.
%%%  These should hopefully be reduced to a single rule, but I can't see
%%%  how to do that elegantly at the moment.

urule(pushUACT1, ruletype3,
   [ in(latest_moves,Mv),
     Mv::val#rec(atype,Move),
     Move::val#rec(pred,assert),
     val(hearer,DP),
     val#rec(wˆcduˆid,IDname) ],
   [ push#rec(wˆgndˆobl,record([pred=uact,
```

```
                                dp=DP,
              args=stackset([IDname])]))]
     ).

urule(pushUACT2, ruletype3,
   [ in(latest_moves,Mv),
     Mv::val#rec(atype,Move),
     Move::val#rec(pred,inforeq),
     val(hearer,DP),
     val#rec(w^cdu^id,IDname) ],
   [ push#rec(w^gnd^obl,record([pred=uact,
                                dp=DP,
              args=stackset([IDname]))])]
     ).

urule(pushUACT3, ruletype3,
   [ in(latest_moves,Mv),
     Mv::val#rec(atype,Move),
     Move::val#rec(pred,agree),
     val(hearer,DP),
     val#rec(w^cdu^id,IDname) ],
   [ push#rec(w^gnd^obl,record([pred=uact,
                                dp=DP,
              args=stackset([IDname]))])]
     ).

urule(pushUACT4, ruletype3,
   [ in(latest_moves,Mv),
     Mv::val#rec(atype,Move),
     Move::val#rec(pred,check),
     val(hearer,DP),
     val#rec(w^cdu^id,IDname) ],
   [ push#rec(w^gnd^obl,record([pred=uact,
```

```
                                dp=DP,
                args=stackset([IDname])]))]
   ).

%%%  The next rule handles an info request - the inforeq is pushed
%%%  into the DH in CDU and the obligation to address it into OBL.

%%%    act:      ID:2, info_request(DP,Q)
%%%    effect:   push(OBL,address(o(DP),ID))

urule( doInfoReq, ruletype3,
   [ val(hearer,DP),
     in(latest_moves,Mv),
     Mv::val#rec(atype,Move),
     Move :: val#rec(pred,inforeq),
     val(next_dh_id,HID),
     print_util::next_dh_idname(HID,HIDname) ],
   [ increase(next_dh_id),
     increase(update_cycles),
     push#rec(w^cdu^tognd^dh,Mv),
     Mv::set#rec(id,HIDname),
     push#rec(w^cdu^tognd^obl,record([pred=address,
                                      dp=DP,
                                      args=stackset([HIDname])])) ]
     ).


%%%  Handle asserts; there are two assert rules here, depending on whether
%%%  the assertion has a confidence level of 1 or 2.  As \cite{MPT2000}
%%%  states, the latter will result in the speaker having an appropriate
%%%  commitment in SCP, but not the former.   Note that both asserts need
%%%  to precede the answer rules in order for the latter to work properly.
%%%  Neither asserts nor answers are great in this implementation; asserts
```

```
%%%  are assumed to contain propositions, but it is further assumed that
%%%  the latter are actually partial - they have arguments but no pred
%%%  value, so we get something like:
%%%  assert(DP,prop(pred=X,args=malvern))
%%%  The pred value is completed by the answer act, which finds the
%%%  appropriate info in the question in PDU.

%%%  act:     ID:2, assert(DP,PROP)
%%%  effect:  push(SCP,scp(DP,PROP))
%%%  effect:  push(COND,accept(o(DP),ID) -> scp(o(DP),PROP))

urule( doAssert, ruletype3,
  [ val(latest_speaker,S),
    val(hearer,H),
    in(latest_moves,Mv),
    Mv::val#rec(atype,Move),
    Move :: val#rec(pred,assert),
    Move :: val#rec(prop,PROP),
    Mv :: val#rec(clevel,2),
    val(next_dh_id,HID),
    print_util::next_dh_idname(HID,HIDname) ],
 [  increase(update_cycles),
    increase(next_dh_id),
    push#rec(w^cdu^tognd^dh,Mv),
    Mv::set#rec(id,HIDname),
    push#rec(w^cdu^tognd^scp,record([pred=scp,
                                     dp=S,
                                     prop=PROP])),
    push#rec(w^cdu^tognd^cond,
                        record([atype1=record([pred=accept_ass,
                                               dp=H,
                                               args=stackset([HIDname])]),
                                atype2=record([pred=scp,
```

```
                                                dp=H,
                                                prop=PROP])])),
    push#rec(w^cdu^tognd^obl,record([pred=address_ass,
                                     dp=H,
                                     args=stackset([HIDname])])) ] ).


%%%  act:     ID:1, assert(DP,PROP)
%%%  effect:  push(COND,accept(o(DP),ID) -> scp(o(DP),PROP))

urule( doAssert2, ruletype3,
  [ val(hearer,H),
    in(latest_moves,Mv),
    Mv::val#rec(atype,Move),
    Move :: val#rec(pred,assert),
    Move::val#rec(prop,PROP),
    Mv::val#rec(clevel,1),
    val(next_dh_id,HID),
    print_util::next_dh_idname(HID,HIDname) ],
  [ increase(update_cycles),
    increase(next_dh_id),
    push#rec(w^cdu^tognd^dh,Mv),
    Mv::set#rec(id,HIDname),
    push#rec(w^cdu^tognd^cond,
                      record([atype1=record([pred=accept_ass,
                                             dp=H,
                                             args=stackset([HIDname])]),
                              atype2=record([pred=scp,
                                             dp=H,
                                             prop=PROP])])),
    push#rec(w^cdu^tognd^obl,record([pred=address_ass,
                                     dp=H,
                 args=stackset([HIDname])])) ]
```

```
  ).

%%%  Handle an answer to a clevel 2 assertion - note the faffing to pick
%%%  up the identity of the act that's being answered

%%%  act:     ID:2, answer(DP,ID2,ID3)
%%%  effect:  push(SCP,ans(DP,Q(ID1),P(ID2)))

urule( doAnswer, ruletype3,
  [ in(latest_moves,Mv),
    Mv::val#rec(atype,Move),
    Move :: val#rec(pred,answer),
    in#rec(w^pdu^tognd^dh,ACT),
    ACT :: val#rec(atype,ATYPE),
    ATYPE :: val#rec(pred,inforeq),
    ACT :: val#rec(id,ACTID),
    val(next_dh_id,HID),
    print_util::next_dh_idname(HID,HIDname) ],
  [ increase(update_cycles),
    increase(next_dh_id),
    push#rec(w^cdu^tognd^dh,record([atype=record([pred=answer,
                                                  dp=c,
                                                  args=stackset([ACTID])]),
                                    clevel=2,
                                    id=HIDname ]))]
  ).

%%%  Deal with a "direct" act.  The act itself is recorded in DH in CDU,
%%%  an obligation to address it is pushed into OBL in CDU, and a
%%%  conditional is added to COND in CDU stating that if the hearer
%%%  accepts the directive then there will be an obligation to carry
%%%  out the instruction.
```

```
%%%  act:     ID:2, direct(DP,Act)
%%%  effect:  push(OBL,address(o(DP), ID))
%%%  effect:  push(COND,accept(o(DP),ID) -> obl(o(DP),Act))

urule( doDirect, ruletype3,
     [ val(hearer,H),
       in(latest_moves,Mv),
       Mv::val#rec(atype,Move),
       Move :: val#rec(pred,direct),
       Move :: val#rec(atype,ATYPE),
       val(next_dh_id,HID),
       print_util::next_dh_idname(HID,HIDname) ],
     [ increase(update_cycles),
       increase(next_dh_id),
       push#rec(w^cdu^tognd^dh,Mv),
       Mv::set#rec(id,HIDname),
       push#rec(w^cdu^tognd^obl,record([pred=address,
                                        dp=H,
                    args=stackset([HIDname])])),
       push#rec(w^cdu^tognd^cond,
                        record([atype1=record([pred=accept,
                                               dp=H,
                                                       args=stackset([HIDname])]),
                                atype2=record([pred=obl,
                                               dp=H,
                                               atype=ATYPE])])) ]
    ).

%%%  Handle a "check".  Record the act, push an obligation for the hearer
%%%  to address the check, and push a conditional to say that if the
%%%  hearer agrees with the check then the checker will be committed to the
%%%  relevant proposition.
```

```
%%%  act:     ID:2, check(DP,PROP)
%%%  effect:  push(OBL,address(o(DP),ID))
%%%  effect:  push(COND,agree(o(DP)),ID -> scp(DP,PROP))

urule( doCheck, ruletype3,
  [ val(hearer,H),
    val(latest_speaker,S),
    in(latest_moves,Mv),
    Mv::val#rec(atype,Move),
    Move :: val#rec(pred,check),
    Move :: val#rec(prop,PROP),
    val(next_dh_id,HID),
    print_util::next_dh_idname(HID,HIDname) ],
  [ increase(update_cycles),
    increase(next_dh_id),
    push#rec(w^cdu^tognd^dh,Mv),
    Mv::set#rec(id,HIDname),
    is::pePathRec(w^cdu^tognd,w^pdu^tognd),
    push#rec(w^cdu^tognd^obl,record([pred=address,
                                     dp=H,
                                     args=stackset([HIDname])])),
    push#rec(w^cdu^tognd^cond,
                      record([atype1=record([pred=agree,
                                             dp=H,
                                             args=stackset([HIDname])]),
                              atype2=record([pred=scp,
                                             dp=S,
                                             prop=PROP])])) ]
   ).

%%%  Handle an "accept" - two rules here because we may have two
%%%  accept moves in latest_moves.  This could be done by altering
%%%  the algorithm, but it also suggests that accepting a
```

```
%%%  directive and accepting an assertion are different kinds of act.
%%%  The first rule does the directs and the second the assertions

urule( doAccept, ruletype3,
  [ in(latest_moves,Mv),
    Mv::val#rec(atype,Move),
    Move :: val#rec(pred,accept),
    val(next_dh_id,HID),
    print_util::next_dh_idname(HID,HIDname) ],
  [ increase(update_cycles),
    increase(next_dh_id),
    Mv::set#rec(id,HIDname),
    push#rec(w^cdu^tognd^dh,Mv) ]
  ).

urule( doAccept2, ruletype3,
  [ in(latest_moves,Mv),
    Mv::val#rec(atype,Move),
    Move :: val#rec(pred,accept_ass),
    val(next_dh_id,HID),
    print_util::next_dh_idname(HID,HIDname) ],
  [ increase(update_cycles),
    increase(next_dh_id),
    Mv::set#rec(id,HIDname),
    push#rec(w^cdu^tognd^dh,Mv) ]
  ).

%%%  Process an "agree" move - as with answers and uacts, we need to
%%%  identify the act that is being agreed with, hence the complications.
%%%  act:     ID:2, agree(DP,ID2)
%%%  effect:  push(SCP,scp(DP,P(ID2)))

urule( doAgree, ruletype3,
```

```
  [ val(latest_speaker,S),
    in(latest_moves,Mv),
    Mv::val#rec(atype,Move),
    Move :: val#rec(pred,agree),
    in#rec(w^pdu^tognd^dh,ACT),
    ACT :: val#rec(atype,ATYPE),
    ATYPE :: val#rec(pred,check),
    ATYPE :: val#rec(prop,PROP),
    ACT :: val#rec(id,ACTID),
    val(next_dh_id,HID),
    print_util::next_dh_idname(HID,HIDname) ],
  [ increase(update_cycles),
    increase(next_dh_id),
    push#rec(w^cdu^tognd^dh,record([atype=record([pred=agree,
                                                  dp=S,
                                                  args=stackset([ACTID])]),
                                    clevel=2,
                                    id=HIDname])),
    push#rec(w^cdu^tognd^scp,record([pred=scp,
                                     dp=S,
                                     prop=PROP])) ]
  ).

%%%  Determine whether a "but" turn conveys concession and update IS with
%%%  move type (i.e., rhetorical relation). (Kavita, 11/6/03).
%%%  (following roughly the doDofe urule template.)

urule( test_concessive_But, ruletype3,
  [ val(latest_speaker,S),
    in(latest_moves,Mv),
    Mv::val#rec(atype,Move),
%access previous turn record
    in#rec(w^pdu^tognd^dh,ACT),
```

```
    ACT::val#rec(atype,ATYPE),
    ACT::val#rec(id,ACTID),
%get proposition and arguments
    ATYPE::val#rec(prop,APROP),
    APROP::val#rec(pred,APRED),
%get CDU assertion CA
    in#rec(w^cdu^tognd^dh,CURACT),
    CURACT::val#rec(atype,CURATYPE),
    CURACT::val#rec(id,CURACTID),
%get proposition and arguments
    CURATYPE::val#rec(prop,CURPROP),
    CURPROP::val#rec(pred,CURPROPPRED),
%get current ID
   val(next_dh_id,HID),
   print_util::next_dh_idname(HID,HIDname),
   test_concession::concession(APRED,CURPROPPRED,ACTID,CURACTID,ConArgs)],
  [ increase(update_cycles),
    increase(next_dh_id),
    push#rec(w^cdu^tognd^dh,record([atype=record([pred=concession
dp=S,
args=stackset([ACTID,
                                                    CURACTID,ConArgs])]),
                  clevel=2,
                id=HIDname])),
    push#rec(w^cdu^tognd^cond,
                     record([atype1=record([pred=accept,
                                            dp=H,
                                            args=stackset([HIDname])]),
                             atype2=record([pred=scp,
                                            dp=H,
      prop=record([pred=concession,
                                                          args=stackset(
                                                            [ConArgs])])])])])
```

```
%%%  --------------          RULETYPE 4          --------------------
%%%  (cf D1.1 p43).  Handle obligation resolution.
%%%  There are three rules here - one is explicitly looking for "address"
%%%  obligations and the presence of a "check" and an appropriate "agree"
%%%  in the DH (cf. the last rule on p37).  The second rule is the more
%%%  general check, and the basic format is simple -  if there's an
%%%  actiontype in OBL and an identical value for atype in DH, then the
%%%  obligation can be removed.  The third handles cases where the need
%%%  to address an assertion is removed by an appropriate check.
%%%  The actual situation is complicated by the type system, as the notes
%%%  on the "obligationRes" rule suggest.
%%%  If there's a "check" and an appropriate "agree", this constitutes an
%%%  "address"  (cf. the last rule on D1.1 p37), which triggers
%%%  the resolution of an obligation.

urule( checkAgree, ruletype4,
  [ in#rec(w^gnd^obl,OBL),
    OBL::val#rec(pred,address),
    OBL::val#rec(dp,DP),
    OBL::val#rec(args,stackset([ACT])),
    in#rec(w^gnd^dh,DACT),
    DACT::val#rec(atype,ATYPE),
    ATYPE::val#rec(pred,agree),
    ATYPE::val#rec(dp,DP),
    ATYPE::val#rec(args,stackset([ACT])),
    in#rec(w^gnd^dh,DACT2),
    DACT2::val#rec(atype,ATYPE2),
    ATYPE2::val#rec(pred,check),
    DACT2::val#rec(id,ACT) ],
  [ delRec(w^gnd^obl,OBL) ]
    ).
```

```
%%%  The main check has to look at the predicate types to see if the records
%%%  match.  The ptypes resource does the checking.

urule( obligationRes, ruletype4,
  [ in#rec(w^gnd^obl,ATYPE),
    in#rec(w^gnd^dh,ACTION),
    ACTION :: val#rec(atype,ATYPE2),
    ATYPE2 :: val#rec(pred,PRED),
    ptypes::ptype(PRED,PREDTYPE),
    ATYPE :: val#rec(pred,PREDTYPE),
    ATYPE :: val#rec(dp,DP),
    ATYPE2 :: val#rec(dp,DP),
    ATYPE :: fst#rec(args,ARG1),
    ATYPE2 :: fst#rec(args,ARG1) ],
  [ delRec(w^gnd^obl,ATYPE) ]
  ).

%%%  This rule looks to see if there's a check that discharges an
%%%  obligation to perform an understanding act.

urule( uactAndCheck, ruletype4,
  [ in#rec(w^gnd^obl,ATYPE1),
    ATYPE1 :: val#rec(pred,uact),
    ATYPE1 :: val#rec(dp,DP),
    ATYPE1 :: fst#rec(args,ARG),
    in#rec(w^gnd^dh,ACTION),
    ACTION :: val#rec(atype,ATYPE2),
    ATYPE2 :: val#rec(dp,DP),
    ATYPE2 :: val#rec(duid,ARG) ],
  [ delRec(w^gnd^obl,ATYPE1),
    delRec(w^udus,ARG) ]
  ).
```

```
%%%  If there's an obligation to address an assertion and a check of the
%%%  assertion in question, then the obligation can be removed.

urule(check_assert, ruletype4,
   [ in#rec(w^gnd^obl,ATYPE),
     ATYPE::val#rec(pred,address_ass),
     ATYPE::val#rec(dp,DP),
     ATYPE::val#rec(args,stackset([ACT])),
     in#rec(w^gnd^dh,DACT),
     DACT::val#rec(id,ACT),
     DACT::val#rec(atype,ATYPE1),
     ATYPE1::val#rec(prop,PROP),
     in#rec(w^gnd^dh,ACT2),
     ACT2::val#rec(atype,ATYPE2),
     ATYPE2::val#rec(dp,DP),
     ATYPE2::val#rec(prop,PROP) ],
   [ del#rec(w^gnd^obl,ATYPE) ]
  ).

%%%  ---------------          RULETYPE 5          --------------------
%%%  (cf D1.1 p43).  Do implicational acts
%%%  The first rule deals with "accept -> obliged" cases, in which
%%%  acceptance of the first act results in an obligation to perform
%%%  the second.

urule( acceptOblW, ruletype5,
  [ in#rec(w^gnd^cond,COND),
    COND::val#rec(atype1,ATYPE1),
    ATYPE1::val#rec(pred,accept),
    COND::val#rec(atype2,ATYPE2),
    ATYPE2::val#rec(pred,obl),
    in#rec(w^gnd^dh,DACT),
    DACT::val#rec(atype,ATYPE1) ],
```

```
  [ increase(update_cycles),
    push#rec(w^gnd^obl,ATYPE2),
    delRec(w^gnd^cond,COND) ]
  ).

%%%  This rule handles "accept -> scp" cases, in which
%%%  acceptance of the first act results in a commitment to
%%%  the second.

urule( accept_scp, ruletype5,
  [ in#rec(w^gnd^cond,COND),
    COND::val#rec(atype1,ATYPE1),
    ATYPE1::val#rec(pred,accept_ass),
    COND::val#rec(atype2,ATYPE2),
    ATYPE2::val#rec(pred,scp),
    in#rec(w^gnd^dh,DACT),
    DACT::val#rec(atype,ATYPE1) ],
  [ increase(update_cycles),
    push#rec(w^gnd^scp,ATYPE2),
    delRec(w^gnd^cond,COND) ]
  ).

%%%  This rule handles "agree -> scp" cases, in which
%%%  agreement with the first act results in a commitment to
%%%  the second.

urule( agree_scp, ruletype5,
  [ in#rec(w^gnd^cond,COND),
    COND::val#rec(atype1,ATYPE1),
    ATYPE1::val#rec(pred,agree),
    COND::val#rec(atype2,ATYPE2),
    ATYPE2::val#rec(pred,scp),
    in#rec(w^gnd^dh,DACT),
```

```
    DACT::val#rec(atype,ATYPE1) ],
  [ increase(update_cycles),
    push#rec(w^gnd^scp,ATYPE2),
    delRec(w^gnd^cond,COND) ]
  ).

%%%  Now the "accept_ass - > scp" cases.  This is complicated because
%%%  there isn't actually an "accept_ass" act.  Instead, if there's a
%%%  check on the assertion in question, and an agree with the check,
%%%  then that constitutes acceptance.

urule( accept_ass_scp, ruletype5,
  [ in#rec(w^gnd^cond,COND),
    COND::val#rec(atype1,AT1),
    AT1::val#rec(pred,accept_ass),
    COND::val#rec(atype2,AT2),
    AT2::val#rec(pred,scp),
    AT1::val#rec(dp,DP),
    AT2::val#rec(prop,PROP),
    in#rec(w^gnd^dh,ACT),
    ACT::val#rec(atype,ATYPE),
    ATYPE::val#rec(pred,check),
    ATYPE::val#rec(dp,DP),
    ATYPE::val#rec(prop,PROP),
    ACT::val#rec(id,CheckID),
    in#rec(w^gnd^dh,ACT2),
    ACT2::val#rec(atype,ATYPE2),
    ATYPE2::val#rec(pred,agree),
    ATYPE2::val#rec(args,stackset([CheckID])) ],
  [ increase(update_cycles),
    push#rec(w^gnd^scp,AT2),
    delRec(w^gnd^cond,COND) ]
  ).
```

```
%%%  Finally, quit.

urule( quit, ruletype5,
   [ val(latest_speaker,c),
     in(latest_moves,Mv),
     Mv::val#rec(atype,Move),
     Move :: val#rec(pred,assert),
     Move :: val#rec(prop,PROP),
     PROP :: val#rec(pred,quit) ],
   [ set( program_state, quit ) ]
  ).

%%%  ---------------          RULETYPE 6          --------------------
%%%  Intention-handling rules (cf. p44)
%%%  This looks to see if the first obligation is to perform
%%%  an understanding act, and if so, pushes the intention to acknowledge
%%%  the act in question. (cf. the first rule on D1.1., p44)

urule( pushAck, ruletype6,
  [ val(latest_speaker,c),
    fstRec(w^gnd^obl,OBL),
    OBL :: val#rec(dp,w),
    OBL :: val#rec(pred,uact),
    OBL :: val#rec(args,ARGS),
    not(in#rec(w^cdu^tognd^dh,ACT) and
       (ACT::val#rec(clevel,1))) ],
  [ push#rec(w^int, record([pred=ack,
                            dp=w,
                            args=ARGS])) ]
  ).

%%%  Push the intention to accept a directive, based on the obligation
```

```
%%%  to address it:

urule( pushAcceptDirect, ruletype6,
  [ val(latest_speaker,c),
    in#rec(w^cdu^tognd^obl,OBL),
    OBL :: val#rec(dp,w),
    OBL :: val#rec(pred,address),
    OBL :: val#rec(args, stackset([CA])),
    in#rec(w^cdu^tognd^dh,ACT),
    ACT :: val#rec(atype,ATYPE),
    ATYPE :: val#rec(pred,direct),
    ACT :: val#rec(id,CA) ],
  [ push#rec(w^int, record([pred=accept,
                            dp=w,
                            args=stackset([CA])])) ]
  ).

%%%  Push the intention to accept an assert, based on the obligation
%%%  to address it:

urule( pushAcceptAssert, ruletype6,
  [ val(latest_speaker,c),
    in#rec(w^cdu^tognd^obl,OBL),
    OBL :: val#rec(dp,w),
    OBL :: val#rec(pred,address_ass),
    OBL :: val#rec(args, stackset([CA])),
    in#rec(w^cdu^tognd^dh,ACT),
    ACT :: val#rec(atype,ATYPE),
    ACT :: val#rec(clevel,2),
    ATYPE :: val#rec(pred,assert),
    ACT :: val#rec(id,CA) ],
  [ push#rec(w^int, record([pred=accept_ass,
                            dp=w,
```

```
                          args=stackset([CA])])) ]
  ).

%%%  A version of David's planning/anticipation rule, albeit a very specific
%%%  one, which produces intentions by checking CDU for the implications of
%%%  other intentions.

urule( checkImplInts, ruletype6,
   [ val(latest_speaker,c),
     in#rec(w^cdu^tognd^cond,CA),
     CA :: val#rec(atype1,ATYPE1),
     ATYPE1 :: val#rec(pred,accept),
     CA :: val#rec(atype2,ATYPE2),
     ATYPE2 :: val#rec(pred,obl),
     in#rec(w^int,ATYPE1),
     ATYPE2 :: val#rec(atype,ATYPE) ],
   [ push#rec(w^int,ATYPE) ]
   ).

%%%  This rule effectively hands control to the deliberation component,
%%%  and performs some of the the functions of the second rule on p44.
%%%  of D1.1  So, given an intention, see if there are "sub-intentions"
%%%  which need to be performed.
%%%  This is done here by calling a "deliberate" step explicitly - so
%%%  it's rather more all-encompassing, but subsumes the rule in the
%%%  deliverable.

urule( checkSubInts, ruletype6,
   [ val(latest_speaker,c),
     in#rec(w^int,INT),
     INT :: val#rec(pred,PRED),
     deliberate::next_move(PRED,NEWINT) ],
   [ push#rec(w^int,NEWINT) ]
```

```
  ).
```

A.2 Additional Modules

These next outputs and inputs are generated from the records created in *temp_utils.pl* to "can" the system's responses and test the update rules. Basically, when we want to test the program, we enter these inputs which get translated according to the output and input rules below.

```
output_form(record([pred=assert,
                    dp=w,
                    prop=record([pred=considered_yardstick_of_beauty,
                                 args=stackset([gg])])]),
    "Greta Garbo was considered the yardstick of beauty").

output_form(record([pred=assert,
    dp=w,
    prop=record([pred=but_did_marry,
    args=stackset([she])])]),
    "But she did marry").

output_form(record([pred=assert,
    dp=w,
    prop=record([pred=not_necessarily_marry,
 args=stackset([beautiful_people])])]),
 "But beautiful people don't have to marry").

% output forms below for "yes that's odd"--KT
output_form(record([pred=accept_ass,
    dp=w,
    prop=record([pred=yes,
    args=_])]),
    "Yes").
```

```
output_form(record([pred=assert,
                    dp=w,
    prop=record([pred=odd,
    args=stackset([that])])]),
    "that's odd").

%% INPUT FORMS
% added by KT to handle DofE turn-initial "but"

input_form([but,she,never,married],
[record([pred=ack,
         %ack signals understanding and moves stuff in cdu into gnd
         dp=c,
         args=_]),
 record([pred=assert,
 dp=c,
 prop=record([pred=never_married,
                      args=stackset([she])])]),
 record([pred=implicit_accept,
 dp=c,
 args=stackset(['CA2'])]),
%currently this is hacked so that the args agree
record([pred=dofe,
dp=c,
args=_])]).
```

The lines below were added to simulate the system's intentions and generate the appropriate utterance. The output forms match the ones given above.

```
%Greta Garbo was considered the yardstick of beauty.
gg :-  tis:apply_operation(push#rec(w^int,
           record([pred=assert,
           dp=w,
```

```
            prop=record([pred=considered_yardstick_of_beauty,
            args=stackset([gg])])])))).

%But she did marry:
marry :- tis:apply_operation(push#rec(wˆint,
     record([pred=reject_ass,
     dp=w,
             args=_]))),

         tis:apply_operation(push#rec(wˆint,
             record([pred=assert,
                     dp=w,
                     prop=record([pred=but_did_marry,
                     args=stackset([she])])])))).


% But beautiful people don’t have to marry:
denyexp :-  tis:apply_operation(push#rec(wˆint,
  record([pred=assert,
          dp=w,
          prop=record([pred=not_necessarily_marry,
          args=stackset([beautiful_people])])]))),

             tis:apply_operation(push#rec(wˆint,
   record([pred=reject_rule,
           dp=w,
         args=_]))).

%Yes, that’s odd.
yes :-   tis:apply_operation(push#rec(wˆint,
     record([pred=accept_ass,
     dp=w,
     prop=record([pred=yes,
```

```
     args=_])])))),

 tis:apply_operation(push#rec(w^int,
     record([pred=accept_rule,
     dp=w,
     args=stackset(['CA6'])]))).

odd :-   tis:apply_operation(push#rec(w^int,
       record([pred=assert,
       dp=w,
     prop=record([pred=odd,
     args=_])]))).
```

A.2.1 Modelling Concession

This code contains an approach to modelling concession which assumes that both CAs and TAs have interpreted the utterances prior to the code being called. We chose to focus modularly on how concession could be modelled and the examples below can be used to test the concessive update rule called above. However the concessive algorithm described at the end of Chapter 6 has yet to be implemented, so the update rule does not do anything at present. The point of presenting this code is simply to illustrate the sorts of examples we hope to model initially.

```
%concessivedialogue acts list: (but, ... is a hack to be triggered by "but")
acts_list([assert(s,welcome_to_atis),assert(u,hello_i_want_to_fly_from_
boston_to_chicago_on_the_seventeenth_of_june_in_the_morning),
assert(s,ok_but,that_would_mean_you_would_need_to_stop_over_in_nyc),
assert(u,ok)]).

%% call "na" with the right Prolog form:
next_acts(stack([]),Acts) :- acts_list([Acts|_]).
next_acts(stack(DH),Acts) :- na(DH,Acts).
```

```
na([Act|_],Acts) :- acts_list(List),
                    next_act(Act,List,Acts).
na([_|Rest],Acts) :- na(Rest,Acts).
next_act(Act,[Act|[Acts|_]],Acts).
next_act(Act,[_|List],Acts) :- next_act(Act,List,Acts).

% concessive example:
%S: "I'd like a flight from boston to chicago on the 17th of june
%in the morning."
%U: "OK, but that would mean you'd need to stop over in nyc."

next_uacts(stack([assert(s,i_want_to_fly_from_boston_to_chicago_
on_the_seventeenth_of_june_in_the_morning)])).
next_uacts(stack([assert(u,ok_but,that_would_mean_you_would_need_
to_stop_over_in_nyc)])).
```

Here is the concessive update rule which appeared above. It simply follows the same template as the DofE update rule. It appears that the update rules for cross-turn discourse relations can quite easily follow this same template barring only differences in social commitments and obligations, as these will differ for the various relations.

```
%%%  Determine whether a "but" turn conveys concession and
%%%  update IS with move type (i.e., discourse relation).

urule( test_concessive_But, ruletype3,
  [ val(latest_speaker,S),
    in(latest_moves,Mv),
    Mv::val#rec(atype,Move),
%access previous turn record
    in#rec(w^pdu^tognd^dh,ACT),
    ACT::val#rec(atype,ATYPE),
    ACT::val#rec(id,ACTID),
%get proposition and arguments
    ATYPE::val#rec(prop,APROP),
```

```
    APROP::val#rec(pred,APRED),
%get CDU assertion CA
    in#rec(w^cdu^tognd^dh,CURACT),
    CURACT::val#rec(atype,CURATYPE),
    CURACT::val#rec(id,CURACTID),
%get proposition and arguments
    CURATYPE::val#rec(prop,CURPROP),
    CURPROP::val#rec(pred,CURPROPPRED),
%get current ID
    val(next_dh_id,HID),
    print_util::next_dh_idname(HID,HIDname),
    test_concession::concession(APRED,CURPROPPRED,ACTID,CURACTID,ConArgs)],
  [ increase(update_cycles),
    increase(next_dh_id),
    push#rec(w^cdu^tognd^dh,record([atype=record([pred=concession,
          dp=S,
  args=stackset([ACTID,
                                                    CURACTID,ConArgs])]),
                    clevel=2,
                  id=HIDname])),
 push#rec(w^cdu^tognd^cond,
                      record([atype1=record([pred=accept_rule,
                                             dp=H,
                                             args=stackset([HIDname])]),
                                 atype2=record([pred=scp,
                                             dp=H,
      prop=record([pred=concession,
                                             args=stackset([ConArgs])])])])])])
```

The above predicate, although not implemented yet, will test whether concession holds. The arguments APRED and CURPROPPRED hold the propositions in the previous and current turns, with ACTID and CURACTID holding the DUs' respective IDs. ConArgs is a list which will be returned by the *test_concession* predicate with the TC and speakers' relations w.r.t. this TC (i.e., X and Y according to the concessive

scheme outlined in Chapter 6).

We predict the concessive relation holding across the ”but” via a TC equivalent to U’s goal in this case, which is $get[flight(from[Boston], to[chicago], on[17\,june, morning])]$.

Determining concessive expectations is trivial in this case, with U arguing for the TC (i.e., the goal succeeds) and S countering the TC (and predicts that the goal cannot succeed as initially planned–i.e., B introduces a hitch in the plan).

Determining the TC in TOD requires a TA interpreter which accesses the speakers’ task-plans. In this case the TC is contained in U’s turn explicitly and is her goal. We return as arguments the TC, the pro-TC element and the con-TC element, which in this case are trivially the utterances themselves. We also pass back the IDs of the turns that are pro and con so that the arguments and speakers (and therefore their argumentative orientation) can be sourced, as well as enabling the arguments themselves to be recovered later on without storing a whole bunch of redundant information. So ConArgs will hold (given a successful *test_concession* call) a list containing $[TC, pro[proposition], against[proposition]$.

Bibliography

Allen, J. and Core, M. (1996). DAMSL: Dialog act markup in several layers. Technical report, University of Rochester.

Allen, J. and Perrault, C. (1980). Analyzing intention in utterances. *Artificial Intelligence*, **15**, 143–178.

Asher, N. (1993). *Reference to Abstract Objects*. Kluwer Academic Publishers.

Asher, N. (1998). Common ground, corrections and coordination. *Journal of Semantics*.

Asher, N. and Lascarides, A. (1998a). Bridging. *Journal of Semantics*, **15**, 83–113.

Asher, N. and Lascarides, A. (1998b). Questions in dialogue. *Linguistics and Philosophy*, **23:2**.

Asher, N. and Lascarides, A. (2001). Indirect speech acts. *Synthese*, **128**, 183–228.

Austin, J. (1962). *How To Do Things With Words*. Clarendon Press, Oxford.

Barwise, J. and Perry, J. (1983). *Situations and Attitudes*. MIT Press.

Blakemore, D. (1987). *Semantic Constraints on Relevance*. Blackwell, Oxford.

Blaylock, N. and Allan, D. (2005). A collaborative problem-solving model of dialogue. *Proceedings of the 6th SIGdial Workshop on Discourse and Dialogue*.

Carberry, S., Chu-Carroll, J., and Green, N. (1993). Rhetorical relations: Necessary but not sufficient. *Association for Computational Linguistics*.

Carletta, J., Isard, A., Isard, S., Kowtko, J., Doherty-Sneddon, G., and Anderson, A. (1996). Hcrc dialogue structure coding manual. Technical report hcrc tr-82, Human Communication Research Centre, University of Edinburgh.

Carlson, L. (1985). *Dialogue Games*. D.Reidel Publishing Company.

Clark, H. (1992). *Arenas of Language Use*. University of Chicago Press.

Clark, H. (1996). *Using Language*. Cambridge University Press.

Cohen, P. and Levesque, R. (1990). Performatives in a rationally based speech act theory. *Association for Computational Linguistics*.

Cooper, R. and Larsson, S. (1999). Dialogue moves and information states. *Proceedings of the Third International Workshop on Computational Semantics*.

Di Eugenio, B., Jordan, P., and Pylkkanen, L. (1998). The coconut project: Dialogue annotation manual. Technical report 98-1, Intelligent Systems Program, University of Pittsburgh.

Dinesh, N., Lee, A., Miltsakaki, E., Prasad, R., Joshi, A., and Webber, B. (2005). Attribution and the (non)-alignment of syntactic and discourse arguments of connectives. *Proceedings of the ACL Workshop on Frontiers in Corpus Annotation II: Pie in the Sky*.

Discourse Research Initiative (1997). Standards for dialogue coding in natural language processing. Technical report, Report no. 167, Dagstuhl Seminar.

Ginzburg, J. (1995a). Resolving questions, part 1. *Linguistics and Philosophy*, **18**, 459–527.

Ginzburg, J. (1995b). Resolving questions, part 2. *Linguistics and Philosophy*, **18**, 459–527.

Ginzburg, J. (1996). Interrogatives: Questions, facts and dialogue. In S. Lappin, editor, *The Handbook of Contemporary Semantic Theory*. Blackwell, Oxford, U.K.

Ginzburg, J. (1997). Querying and assertions in dialogue. *Draft paper*.

Ginzburg, J. (1998). Clarifying utterances. *Proceedings of the Twente Workshop on the Formal Semantics and Pragmatics of Dialogue*.

Goffman, E. (1976). Replies and responses. *Language and Society*, **5**, 257–313.

Graesser, A., Person, N., and Huber, J. (1992). Mechanisms that generate questions. In T. Lauer, E. Peacock, and A. Graesser, editors, *Questions and Information Systems*. Erlbaum, Hillsdale, NJ.

Green, N. and Carberry, S. (1992). Conversational implicatures in indirect replies. *Association for Computational Linguistics*.

Grice, H. (1975). Logic and conversation. In P. Cole and J. Morgan, editors, *Speech Acts*.

Grosz, B. and Kraus, S. (1999). The evolution of sharedplans. In A. Rao and M. Wooldridge, editors, *Foundations and Theories of Rational Agencies*.

Grosz, B. and Sidner, C. (1986). Attention, intentions and the structure of discourse. *Computational Linguistics*, **12**, 175–204.

Grosz, B., Joshi, A., and Weinstein, S. (1995). Centering: A framework for modelling the local coherence of discourse. Technical report 95-01, IRCS, University of Pennsylvania.

Grote, B., Lenke, N., and Stede, M. (1995). Ma(r)king concessions in english and german. *In Proceedings of the Fifth European Workshop on Natural Language Generation*.

Heeman, P. (2004). *Speech Repairs, Intonational Boundaries and Discourse Markers in Spoken Dialog*. Ph.D. thesis, Computer Science Department, University of Rochester.

Hobbs, J. (1985). On the coherence and structure of discourse. Report no. csli-85-37, CSLI, Stanford University.

Hobbs, J. (1996). On the relation between informational and intentional perspectives on discourse. In E. Hovy and D. Scott, editors, *Burning Issues in Discourse: An Inter-Disciplinary Account, volume 151 of NATO ASI Series, Series F: Computer and Systems Sciences*. Springer-Verlag, Berlin.

Hovy, E. (1993). Automated discourse generation using discourse structure relations. *Artificial Intelligence, Special Issue on NLP*, **63**.

Jackendoff, R. (1972). *Semantic Interpretation in Generative Grammar*. MIT Press.

Kamp, H. and Reyle, U. (1993). *From Discourse to Logic: Introduction to ModelTheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. Kluwer Academic Publishers, London, Boston, Dordrecht.

Karagjosova, E. (2001). Towards a comprehensive meaning of german 'doch'. *European Summer School for Logic, Language and Information (ESSLLI) Student Session, Helsinki*.

Knott, A. (1996). *A Data-Driven Methodology for Motivating a Set of Coherence Relations*. Ph.D. thesis, Department of Artificial Intelligence, University of Edinburgh.

Knott, A. (1999a). An algorithmic framework for specifying the semantics of discourse relations. *Proceedings of the Pacific Association for Computational Linguistics*.

Knott, A. (1999b). Discourse relations as descriptions of an algorithm for perception, action and theorem-proving. *In Proceedings of the International Workshop on Levels of Representation in Discourse (LORID '99)*.

Kreutel, J. and Matheson, C. (1999). Modelling questions and assertions in dialogue using obligations. *Proceedings of Amstelogue 99, 3rd Workshop on the Semantics and Pragmatics of Dialogues, Amsterdam*.

Kreutel, J. and Matheson, C. (2000). Information states, obligations and intentional structure in dialogue modelling. *In Proceedings of the 3rd International Workshop on Human-Computer Conversation*.

Kreutel, J. and Matheson, C. (2001a). Context-dependent interpretation and implicit dialogue acts. *Proceedings of Bibdialog, Bielefeld*.

Kreutel, J. and Matheson, C. (2001b). Incremental information state updates in an obligation-driven dialogue model. *Language and Computation*.

Kruijff-Korbayová, I. and Webber, B. (2000). Presuppositional interpretation of causal and additive 'although'. *Poster and talk given at* Making Sense*, Groningen*.

Kruijff-Korbayová, I. and Webber, B. (2001). Concession, implicature and alternative sets. *In Proceedings of the International Workshop on Computational Semantics (IWCS-4)*.

Lagerwerf, L. (1998). *Causal Connectives Have Presuppositions*. Ph.D. thesis, Catholic University of Brabant, Holland Academic Graphics, The Hague, The Netherlands.

Lakoff, R. (1971). If's, and's and but's about conjunction. In Fillmore and Langedoen, editors, *Studies in Linguistic Semantics*. Holt, Rinehart and Winston, Inc.

Lambert, L. and Carberry, S. (1991). A tripartite plan-based model of dialogue. *Association for Computational Linguistics*.

Larsson, S. (2002). *Issue-Based Dialogue Management*. Ph.D. thesis, Department of Linguistics, Gothenburg University.

Lascarides, A. and Asher, N. (1999). Cognitive states, discourse structure and the content of dialogue. *Proceedings of Amstelogue 1999, Amsterdam*.

Lascarides, A. and Asher, N. (2002). Segmented discourse representation theory: Dynamic semantics with discourse structure. In H. Bunt and R. Muskens, editors, *Computing Meaning: Volume 3*. Kluwer Academic Publishers.

Lehnert, W. (1977). *The Process of Question Answering*.

Lesh, N., Rich, C., and Sidner, C. (1999). Using plan recognition in human-computer collaboration. *International Conference on User Modelling*.

Levinson, S. (1983). *Pragmatics*. Cambridge University Press.

Litman, D. and Allen, J. (1987). A plan recognition model for subdialogues in conversation. *Cognitive Science*, **11**.

Lochbaum, K. (1998). A collaborative planning model of intentional structure. *Computational Linguistics*, **24**.

Maier, E. (1996). Context construction as subtask of dialogue processing – the verbmobil case. Technical report, DFKI GmbH Report 134.

Mani, I., Bloedorn, E., and Gates, B. (1998). Using cohesion and coherence models for text summarization. Papers from the 1998 aaai spring symposium, AAAI Press.

Mann, W. and Thompson, S. (1988). Rhetorical structure theory. *Text*, **8**.

Matheson, C., Poesio, M., and Traum, D. (2000). Modelling grounding and discourse obligations using update rules. *Proceedings of the North American Association for Computational Linguistics*.

Modjeska, N., Markert, K., and Nissim, M. (2003). Using the web in machine learning for 'other'-anaphora resolution. *Proceedings of Empirical Methods in Natural Language (EMNLP 2003)*.

Moore, J. and Paris, C. (1994). Planning text for advisory dialogues: Capturing intentional and rhetorical information. *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics*.

Moore, J. and Pollack, M. (1992). A problem for rst: The need for multi-level discourse analysis. *Computational Linguistics*, **18:4**.

Nakatani, C. and Traum, D. (1999). Coding discourse structure in dialogue (version 1.0). Technical report cs-tr-3980/umiacs-tr-99-03, University of Maryland Institute for Advanced Computer Studies.

Nygren, N. (2004). *Resolving Other-Anaphora*. Ph.D. thesis, School for Informatics, University of Edinburgh.

Oberlander, J. and Knott, A. (1996). Issues in cue phrase implicature. *Working Notes for the AAAI Spring Symposium on Computational Implicature*.

Palmer, M. (1990). *Semantic Processing for Finite Domains*. Cambridge University Press.

Penstein-Rose', C. (1995). The structure of multiple-headed negotiations. Technical report, Carnegie Mellon University Technical Report number CMU-PHIL-65.

Poesio, M. (1995). A model of conversation processing based on micro conversational events. *Proceedings of the 17th Annual Meeting of the Cognitive Society*, pages 698–703.

Poesio, M. (1998). Cross-speaker anaphora and dialogue acts. *Proceedings of Workshop on Mutual Knowledge, Common Ground and Public Information, ESSLLI*.

Poesio, M. and Muskens, R. (1997). The dynamics of discourse situations (extended abstract). *Proceedings of the 11th Amsterdam Colloquium, University of Amsterdam, ILLC*.

Poesio, M. and Traum, D. (1997). Conversational actions and discourse situations. *Computational Intelligence*, **13(3)**.

Poesio, M. and Traum, D. (1998). Towards an axiomatization of dialogue acts. *Proceedings of Twente Workshop*.

Rich, C., Lesh, N., and Sidner, C. (2000). Collagen: Applying collaborative discourse theory to human-computer interaction. Technical report, Mitsubishi Electric Research Laboratory TR2000-038.

Sacks, H. and Schegloff, E. (1973). Opening up closings. *Semiotica*, **8**.

Sanders, T., Spooren, W., and Noordman, L. (1992). Towards a taxonomy of coherence relations. *Discourse Processes*, **15**, 1–35.

Searle, J. (1969). Speech acts: An essay in the philosophy of language. Technical report, Cambridge University Press.

Sidner, C. (1994). An artificial discourse language for collaborative negotiation. *AAAI*.

Sikorski, T. and Allen, J. (1997). A scheme for annotating problem solving actions in dialogue. *Working Notes of AAAI Fall Symposium on Communicative Actions in Humans and Machines*.

Spooren, W. (1989). *Some Aspects of the Form and Interpretation of Global Contrastive Coherence Relations*. Ph.D. thesis, Catholic University of Nijmegen, The Netherlands.

Stent, A. (2000). Rhetorical structure in dialogue. *Proceedings of the 2nd Int'l NLG Conference (INLG2000)*.

Stent, A. and Allen, J. (2000). Annotating argumentation acts in spoken dialog. Technical report, University of Rochester, Technical Report 740.

Stone, M. (2000). Towards a computational account of knowledge, action and inference in instructions. *Language and Computation, vol. 0 No. 0*.

Traum, D. (1994). *A Computational Theory of Grounding in Natural Language Conversation*. Ph.D. thesis, Department of Computer Science, University of Rochester.

Traum, D. and Allen, J. (1994). Discourse obligations in dialogue processing. *Proceedings of the Association for Computational Linguistics*.

Traum, D. and Hinkelman, E. (1992). Conversational acts in task-oriented spoken dialogue. *Computational Intelligence special issue: Computational Approaches to Non-Literal Language*, **8**.

Traum, D. and Nakatani, C. (1999). A two-level approach to coding dialogue for discourse structure: Activities of the 1998 dri working group in higher-level structures. *Association for Computational Linguistics Workshop Towards Standards and Tools for Discourse Tagging*.

Traum, D., Bos, J., Cooper, R., Larsson, S., Lewin, I., Matheson, C., and Poesio, M. (1999). A model of dialogue moves and information state revision. Technical report, Technical Report Deliverable D2.1, TRINDI.

TRINDI, C. (2001). The trindi book. Technical report, Universities of Gothenburg, Edinburgh, Saarlandes and SRI Cambridge and Xerox Research Centre Europe.

Tsovaltsi, D. (2001). *Formalising Hinting in Tutorial Dialogues*. Ph.D. thesis, Division of Informatics, University of Edinburgh.